GEORGIA INSTITUTE OF TECHNOLOGY
OFFICE OF CONTRACT ADMINISTRATION
SPONSORED PROJECT INITIATION

no action add'l

Date: July 25, 1977

Project Title: Program Evaluation Methodologies Training

Project No: E-24-520

Project Director: Dr. Thomas B. Clark

Sponsor: Georgia Department of Human Resources

Agreement Period: From July 1, 1977 Until September 30, 1977

Type Agreement: Contract No. 427-9300200; dated June 23, 1977

Amount:
$6,607 GDHR
2,203 GIT (E-24-213)
$8,810

Reports Required: Monthly Statement of Training Activities; Monthly Report of Achievements; Final Report

Sponsor Contact Person (s):

Technical Matters

Mr. John B. Pinka
Title XX Training Coordinator
Georgia Department of Human Resources
618 Ponce de Leon Avenue, N. E., 2nd Floor
Atlanta, Georgia 30308

Contractual Matters
(thru OCA)

NOTE: Continuation of E-24-516 [C/S E-23-212], which terminated June 30, 1977.

Defense Priority Rating: N/A

Assigned to: Industrial and Systems Engineering (School/Laboratory)

COPIES TO:

- Project Director
- Division Chief (EES)
- School/Laboratory Director
- Dean/Director—EES
- Accounting Office
- Procurement Office
- Security Coordinator (OCA)
- Reports Coordinator (OCA)
- Library, Technical Reports Section
- Office of Computing Services
- Director, Physical Plant
- EES Information Office
- Project File (OCA)
- Project Code (GTRI)
- Other _____

CA-3 (3/76)

GEORGIA INSTITUTE OF TECHNOLOGY
OFFICE OF CONTRACT ADMINISTRATION

SPONSORED PROJECT TERMINATION

No action 08 1978

Date: January 12, 1978

Project Title: Program Evaluation Methodologies Training

Project No: E-24-520

Project Director: Dr. Thomas B. Clark

Sponsor: Georgia Department of Human Resources

Effective Termination Date: 9/30/77

Clearance of Accounting Charges: 9/30/77

Grant/Contract Closeout Actions Remaining: NONE

_ Final Invoice and Closing Documents
_ Final Fiscal Report
_ Final Report of Inventions
_ Govt. Property Inventory & Related Certificate
_ Classified Material Certificate
_ Other _____

Assigned to: Industrial & Systems Engineering (School/Laboratory)

COPIES TO:

Project Director
Division Chief (EES)
School/Laboratory Director
Dean/Director–EES
Accounting Office
Procurement Office
Security Coordinator (OCA) ✓
Reports Coordinator (OCA)
Library, Technical Reports Section
Office of Computing Services
Director, Physical Plant
EES Information Office
Project File (OCA)
Project Code (GTRI)
Other _____

CA–4 (3/76)

E-24-520

Georgia Department of Human Resources
Staff Development Section

MONTHLY REPORT OF TRAINING ACTIVITIES

Training Unit Program Evaluation Methodologies Month of August, 19 77

TITLE OF TRAINING ACTIVITY *	LENGTH OF TRAINING SESSION	NUMBER OF TRAINEES PER SESSION	NUMBER OF MAN TRAINING DAYS
Geneen Mills & Pat Tanner	2 hrs.	2	4 man-hrs.
Jerry Legge	8 "	1	8 "
Gloria Peluso	2 "	1	2 "
Marietta Suhart	2 "	1	2 "
Harry Tomas	6 "	1	6 "
*Please refer to items A through E under attached Monthly Report of Achievements	"Objectives Achieved" on the in Title XX Training Activities.		

Sept. 7, 1977
DATE

SIGNATURE AND TITLE

Thomas B. Clark
Assistant Professor - ISyE
Georgia Tech

Send To: Staff Development Section
Room 341-H
47 Trinity Avenue, S.W.
Atlanta, Georgia 30334

See Instructions On Reverse Side

OC(4)-33 (9-76)

Georgia Department of Human Resources
Staff Development Section

MONTHLY REPORT OF TRAINING ACTIVITIES

(9/1 - 9/15)

Training Unit _ Program Evaluation Methodologies Month of September, 19 77

TITLE OF TRAINING ACTIVITY *	LENGTH OF TRAINING SESSION	NUMBER OF TRAINEES PER SESSION	NUMBER OF MAN TRAINING DAYS
Norma Edwards	2 hrs.	1	2 man hrs.
Harry Tomas	10 hrs.	1	10 man hrs.
*Please refer to items A and B under "Objectives Achieved" on the attached Monthly Report of Achievements in Title XX Training Activities			

December 13, 1977
DATE

SIGNATURE AND TITLE

Thomas B. Clark, Ph.D.
Assistant Professor - ISyE
Georgia Tech

Send To: Staff Development Section
Room 341-H
47 Trinity Avenue, S.W.
Atlanta, Georgia 30334

See Instructions On Reverse Side

OC(4)-33 (9-76)

E-24-520

GEORGIA DEPARTMENT OF HUMAN RESOURCES

MONTHLY REPORT OF ACHIEVEMENTS IN TITLE XX TRAINING ACTIVITIES

UBMIT TO: Department of Human Resources
DHR Staff Development Section
Room 341-H
47 Trinity Avenue, S. W.
Atlanta, Georgia 30334

Due by the 20th day of the following month)

For the Month of August 1977

Georgia Tech
Name of Educational Institution

Program Evaluation Methodologies
Program Title

Thomas B. Clark, Program Director

Objectives achieved (Describe in narrative form objectives achieved during the month as they relate to the overall goal set forth in your proposal):

A. Met with Geneen Mills (University of Georgia Center for Continuing Education) and Pat Tanner to discuss needs assessment and evaluation design for staff training.

B. Completed written review of DHR Needs Assessment Model and a published resource allocation methodology for Jerry Legge.

C. Provided technical assistance to Gloria Peluso (Augusta College) in statistical analysis of evaluation test data.

D. Critiqued pre-test instrument for Marietta Suhart (University of Georgia Center for Continuing Education) for homemaker aids training.

E. Continued work with Harry Tomas (Georgia Tech graduate student) in the development of a financial model for the evaluation of sliding fee scales for day care services.

Problems encountered (Describe any problems which may have affected your objectives and/or goals):

None

E-24-520

GEORGIA DEPARTMENT OF HUMAN RESOURCES

MONTHLY REPORT OF ACHIEVEMENTS IN TITLE XX TRAINING ACTIVITIES

(9/1 - 9/15)

BMIT TO: Department of Human Resources
DHR Staff Development Section
Room 341-H
47 Trinity Avenue, S. W.
Atlanta, Georgia 30334

For the Month of September 19 77

Georgia Tech
Name of Educational Institution

Program Evaluation Methodologies
Program Title

ue by the 20th day of the following
onth)

Thomas C. Clark, Program Director

Objectives achieved (Describe in narrative form objectives achieved during the month as they relate to the overall goal set forth in your proposal):

A. Critiqued pre-test instrument for Norma Edwards (University of Georgia Center for Continuing Education) for gerantology training.

B. Continued work with Harry Tomas (Georgia Tech graduate student) in the development and computer programming of a financial model for the evaluation of sliding fee scales for day care services.

Problems encountered (Describe any problems which may have affected your objectives and/or goals):

None.

September 30, 1977

FINAL EVALUATION REPORT

for Title XX Training Contract on

PROGRAM EVALUATION METHODOLOGIES

Prepared and Submitted by
Dr. Thomas B. Clark, Program Director
Assistant Professor
School of Industrial & Systems Engineering
Georgia Institute of Technology

ACKNOWLEDGEMENTS

The Program Director expresses his sincere appreciation to Ms. Pat Tanner of the Training Unit, Contract Services Section, Title XX Administration for her administrative support and her active participation in the training program.

TABLE OF CONTENTS

Appendices:

GOALS AND OBJECTIVES

As stated in the program proposal, the overall goal of the training was to provide CSS personnel the necessary skills at appropriate levels of detail and comprehensiveness that will allow them to perform effectively their respective roles in program evaluation endeavors.

The specific short-range objectives of the training were:

1. To impart the specific knowledge and skills to CSS Staff that is required for the effective design, implementation, and use of evaluation systems for social service programs.
2. To equip the CSS Training Staff in particular with the knowledge and skills necessary to:
   a. Use evaluation information to determine training needs,
   b. Evaluate the effectiveness of training programs that they conduct in-house or purchase from outside training contractors, and
   c. Provide future in-house training for CSS Staff on the subject of evaluation.
3. To assure that CSS Staff is able to impart evaluation skills to service provider agency staff by overseeing training programs conducted by CSS trainers.

The expected long-range results of the training were that it would lead to:

1. The development of evaluation systems for CSS-administered programs that are more effective, less costly, and less likely to cause organizational trauma than previous evaluation systems; which in turn should lead to -
2. More effective management of CSS-administered programs to yield greater results for the resources consumed, and the credible documentation of those results; which in turn should lead to -

3. The future availability of resources to maintain, improve, and expand needed social programs.

POPULATION TRAINED

The participants in the training program represented the following organizational elements of Title XX Administration:

1. Training Unit; Contract Services Section
2. Program Unit; Contract Services Section.
3. Research and Evaluation Unit; Management, Planning and Development Section.

The participants are listed in Table I along with their job title, organizational affiliation, and highest academic degree earned. The list is divided into two groups. Group A consists of participants who completed the entire program, including the comprehensive final examination. Group B contains individuals who attended portions of the training but did not "complete" the program, in that they did not take the final examination.

In addition to DHR Staff, a number of training contractors received less intensive exposure to the training content through:

1. A three-hour summary presentation by the Program Director at the Title XX Training Contractors' Meeting held in Atlanta, March 16-17, 1977. (Approximate attendance: 25).
2. Technical assistance provided by the Program Director upon request to other training contractors with respect to the evaluation methodologies employed for their own training programs.

TABLE I
TRAINING PARTICIPANTS

NAME	JOB TITLE	ORGANIZATION AFFILIATION	DEGREE
	- GROUP A -		
Bartling, Betty J.	Program Rep.	Program Unit	MSW
Baughman, Mary V.	Program Rep.	Program Unit	MSW
Blackford, Penny A.	Consultant	Program Unit	MSW
Cash, D. Lee	Research Assoc.	Res. & Eval.	MGA
Gresham, Janice T.	Program Rep.	Program Unit	MSW
Harris, Jo Anne M.	Training Rep.	Training Unit	MEd
Heard, Geraline	Program Rep.	Program Unit	MEd
Johnstone, Gary D.	Program Rep.	Program Unit	MSW
Jones, Evelyn S.	Program Rep.	Program Unit	MEd
Legge, Jerry	Research Assoc.	Res. & Eval.	Ph.D.
Sampson, Dorothy N.	Consultant	Training Unit	MSW
	- GROUP B -		
Carswell, Earl	Program Rep.	Program Unit	MS
Clankscales, Barbara S.	Program Rep.	Program Unit	MSW
Entwistle, Dan A.	Program Rep.	Program Unit	MA
Geoffrey, Joseph J.	Program Rep.	Program Unit	MSW
Hart, Kenneth E.	Training Rep.	Training Unit	Bach.
Howell, John R.	Program Specialist	Support Unit	MSW
Maddorn, Les C.	Regional Team Mgr.	Program Unit	MSW
Patricio, Alan B.	Program Rep.	Program Unit	Bach.
Wallace, Clarice M.	Program Rep.	Program Unit	MA
Wilson, Marilynn M.	Program Rep.	Program Unit	MA
Woodward, Delores A.	Regional Team Mgr.	Program Unit	MS

NEED ASSESSMENT AND COURSE CONTENT

The major thrust of the proposal was to provide training on impact (i.e., results oriented) evaluation methodologies as contrasted with process (i.e., resource and activity oriented) evaluations. Impact evaluation implies a research problem in which an attempt is made to isolate and measure changes that occur in a given system as a result of a specific treatment applied to that system. At the outset, therefore, it was clear that the training program must contain the following sequential elements of research methodology as applied to social service and training program evaluations:

1. Overall concepts of the impact evaluation process.
2. Criteria definition
3. Experimental design
4. Measurement, surveys, and sampling
5. Statistical data analysis.

Some additional topics were deemed necessary as a result of the Program Director's previous experience as a consultant to an in-house DHR evaluation effort. The topics were:

1. Organizational implications of impact evaluations
2. Computer support for data analysis
3. Managing the evaluation process, including concepts of project management with CPM (Critical Path Methodology).

The above eight topic areas were arranged in logical order and outlined in greater detail in the proposal. The outline was reviewed by Ms. Pat Tanner in her role as Staff Development Training Representative and found to be complete and appropriate.

In addition, a comprehensive pre-test (see Appendix A) was given at the start of the intensive classroom instruction program. This pre-test was used

as an additional input to the needs assessment process (and as a base point for the evaluation of the training program as will be explained later). The primary conclusion with respect to need assessment that resulted from the pre-test was that a large percentage of the participants did not possess the prerequisite technical knowledge that would have been required for the planned depth of coverage in the area of statistical data analysis. It was therefore decided to approach that topic with a less technically sophisticated survey of statistical concepts and the general characteristics, appropriate applications, and limitations of several common statistical techniques. A similar decision was made to limit the technical depth of coverage in the area of computer support for data analysis.

The resulting content of the intensive classroom instruction is outlined in great detail in Appendix B.

In addition to determining the content of the intensive classroom training, it was decided in discussions with Ms. Tanner that small group application projects should be used after the completion of the classroom instruction. Each small group of participants would undertake the process of designing, implementing, and analyzing an impact evaluation for a specific "live" social service or training program. Guidance and technical assistance for each group would be provided by the Program Director. The objective was to allow the participants to integrate the concepts presented in the classroom, and begin to test and develop their skills in real applications. In essence, these projects were designed for participants to "get their feet wet" in conducting impact evaluations in a situation that provided help as needed.

DELIVERY MECHANISMS AND TRAINING MATERIALS

The delivery mechanisms and training materials employed are discussed in four sections corresponding to the four types of training activities conducted:

1. Intensive classroom instruction for State Staff.
2. Small group application projects for State Staff
3. Presentation at the Title XX Training Contractors' Meeting
4. Technical assistance to other training contractors.

Intensive Classroom Instruction

The intensive classroom instruction for State Staff was conducted in nine four-hour sessions held weekly on Tuesday afternoons beginning on November 2, 1976 and ending on January 11, 1977. Two Tuesdays were skipped during the Christmas season. Thus, the training involved 36 contact hours of instruction. Those participants who completed the course (see Group A in Table I presented earlier) were awarded 3.6 CEUs, which are recorded in the Georgia Tech Registrar's Office.

The sessions were held in a conference room at the Georgia Mental Health Institute in Atlanta due to lack of available conference rooms at DHR facilities.

The training was conducted primarily in a lecture/discussion format. Note-taking outline guides were handed out at the start of each major topic area. Participants filled in the details on the handouts as they took notes in class. The sessions were kept as informal as possible, and the trainees were encouraged to participate by raising questions and sharing relevant experiences.

Mini-cases, mini-quizzes, and homework problems were also used where appropriate to provide opportunities to apply concepts and techniques presented in the lectures and to monitor comprehension. The mini-quizzes, in particular, also seemed to provide an added degree of motivation for the participants to "keep-up" with the class.

A complete set of the handout materials, including:

1. A bibliography of selected relevant books
2. Note-taking outline guides
3. Mini-cases
4. Mini-quizzes
5. Homework problems

is contained in Appendix C and is arranged in the order that it was presented to the class.

With the exception of one session, the training was conducted by the Program Director. The one exception occurred on November 9, when the Program Director was out of town. That session was conducted by Dr. Terrence Connolly, Associate Professor of Industrial and Systems Engineering at Georgia Tech. Dr. Connolly is an expert in the area of experimental design, which was the primary topic in that session.

Small Group Application Projects

At the completion of the intensive classroom instruction, most of the participants who had completed that phase of the program formed small groups to pursue application projects. The specific programs to be evaluated and the participants involved in each of these application projects were as follows:

1. Contracted adult day care programs. – Penny Blackford, Barbara Clankscales, and Janice Gresham.
2. In-house staff training programs on administrative policies and procedures. - Pat Tanner.
3. Contracted training for social service workers. - Jerry Legge, Pat Tanner, and Barbara Williamson.
4. Contracted training for child day care teachers. - Jo Anne Harris, Ken Hart, and Dot Sampson.

5. Experimental sliding fee scales for contracted child care services - Lee Cash, Evelyn Jones, and Jerry Legge.

With the exception of the project concerning sliding fee scales, each group was to undertake a process that would involve criteria definition, experimental design, development and application of measurement instruments, and data analysis. The Program Director was to provide guidance and technical assistance as needed. It was determined that Friday was the most convenient day of the week in general for these groups to meet and work on their projects. Consequently, the Program Director reserved all Fridays in his schedule to work with these groups as requested. This did not preclude the possibility of work sessions on other days or the Program Director's participation in such sessions.

Each of these four groups made varying degrees of progress on their projects. Because of conflicting pressures on their time and some difficulties in coordinating schedules within the groups, however, none of these four groups devoted as much time to their projects as had been anticipated. As a result, none of the four has yet produced a complete and comprehensive evaluation for their respective programs, as far as the Program Director can ascertain. Some of the projects are still in progress and may yet reach completion.

This lack of follow-through was a disappointing aspect of the training program from the Program Director's point of view. Further, it illustrates a very real danger with respect to the performance of impact evaluations. It is that such evaluations will not be accomplished unless they carry a high administrative priority, perhaps including specific deadlines. The amount and difficulty of the work required and the possibility of unsupportive results lead to a natural inclination to procrastinate.

The fifth group, which was studying experimental sliding fee scales for child day care services, took an entirely different type of approach. After some discussion, it became apparent that a flexible and dynamic mathematical

model was needed by DHR as a means of projecting the economic consequences of alternative sliding fee scales over future years. As a result, the Program Director involved Mr. Harry Tomas (a masters level graduate student in Industrial and Systems Engineering at Georgia Tech) in the project. As his masters thesis, Mr. Tomas will develop a computerized (FORTRAN) simulation model capable of performing the necessary economic projections using data and assumptions provided by DHR. A complete copy of Mr. Tomas' thesis proposal is contained in Appendix D. This effort will be carried to completion as an unfunded academic project outside the provisions of the training contract. The Program Director will serve as the Thesis Advisor. Completion is anticipated by March, 1978.

Title XX Training Contractors' Meeting

As mentioned earlier, the Program Director made a three-hour summary presentation on program evaluation methodologies at the Title XX Training Contractors' Meeting on March 17, 1977. A complete set of the transparencies used in that presentation is provided in Appendix E.

As part of the presentation, the Program Director gave general feedback to the other contractors concerning the adequacy of the evaluation plans contained within twelve training proposals. These proposals had been submitted by contractors attending the meeting and had been reviewed by the Program Director prior to the meeting. The feedback was given in an aggregate fashion, so that individual proposals were not identified in the meeting. This process, however, provided the Program Director an opportunity to offer his assistance to the other contractors in the design and implementation of their evaluation procedures.

Technical Assistance to Training Contractors

Technical assistance was provided by the Program Director upon request to the following training contractors:

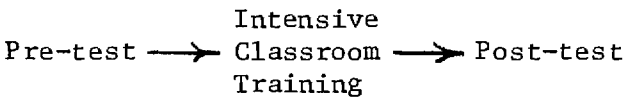
1. Geneen Mills, University of Georgia Center for Continuing Education - State Staff Training
2. Marge Peterson and Gloria Peluso, Augusta College - Day Care Teacher Training.
3. Marietta Suhart, University of Georgia Center for Continuing Education - Homemaker Aide and Gerantology Training.

The assistance was provided via telephone and mail correspondence and through meetings which took place in Atlanta. The assistance related to the following aspects of their evaluation endeavors:

1. Needs assessment
2. Criteria definition
3. Experimental design
4. Development of measurement instruments, primarily tests
5. Sampling plans
6. Statistical data analysis.

EVALUATION

The design employed in the evaluation of the training program was:



The pre-test was mentioned earlier and is contained in Appendix A. The post-test can be found in Appendix F. Both tests are comprehensive, covering concepts presented throughout the course. Both contain short answer questions, discussion questions, and quantitative problems. The tests were administered as closed-notes take-home exams with instructions that each participant was to work independently. Both tests had maximum scores of 100 points.

A number of questions appear on both exams, thus raising the possibility that scores on the post-test may reflect learning with respect to those specific questions, rather than more general learning in the subject area. This possible bias is believed to be minimal, however, for the following two reasons:

1. The pre-tests were not returned to the participants, so it could not have been used as a study guide in preparing for the post-test.
2. There was an eleven week interval between the pre-test and the post-test.

Every effort was made to protect the objectivity of the grading process. Each question had a designated point value, which was printed on the test. Each item, including the discussion questions, required an answer that displayed specific knowledge or skill. Partial credit was given, and a conscious effort was made to grade the post-test more strictly (i.e., less liberal awarding of partial credit) than the pre-test. This policy was intended to avoid creating a spurious increase in test scores from the pre-test to the post-test due to a grading bias.

TABLE II
PRE-TEST AND POST-TEST GRADES

PARTICIPANT NUMBER	PRE-TEST GRADE*	POST-TEST GRADE*	DIFFERENCE (d_i)
1	8.0	76.5	+68.5
2	39.0	88.0	+49.0
3	22.0	93.0	+71.0
4	-	81.0	-
5	9.0	58.5	+49.5
6	14.5	88.5	+74.0
7	12.5	71.0	+58.5
8	30.0	98.5	+68.5
9	74.0	82.0	+ 8.0
10	7.5	80.0	+72.5
11	28.5	89.5	+61.0
12	-	92.0	-
13	40.5	93.5	+53.0

*Maximum grade on each test was 100.

The grades for the thirteen individuals who completed the intensive classroom training are shown in Table II. The grades have not been identified by individual, and the order in which they are presented has been randomized. (Participants number 4 and 12 had not taken the pre-test.)

It was gratifying to see (1) that 12 of the 13 participants who took the post-test scored over 70%, and (2) that every participant who took both the pre-test and the post-test improved his or her score. In order to determine the statistical significance of the increase in scores, two analytical tests were made as described below.

Student t test for differences in paired observations

Let: n = the number of pairs of test scores = 11.

d_i = the differences in individual pairs of test scores = 68.5, 49.0, 71.0, . . . , 53.0.

$\bar{d}$ = the mean of the d_i = $(\sum d_i)/n$ = 6333.5/11 = 57.6

δ = the hypothesized mean of the differences in paired test scores.

Then: H_o: $\delta = 0$ (null hypothesis)

H_a: $\delta > 0$ (alternative hypothesis)

The student t statistic is computed as:

$$t = \frac{(\bar{d}-\delta)(n)^{1/2}}{[\sum (d_i-\bar{d})^2/(n-1)]^{1/2}} \qquad t = \frac{(57.6-0)(11)^{1/2}}{[3541.4/10]^{1/2}}$$

$$t = 10.16$$

with n-1 = 10 degrees of freedom.

The null hypothesis, therefore, is rejected at a level of significance (α) of 0.005.

In addition, a 95% confidence interval for δ was computed as ($45.0 \leq \delta \leq 70.2$).

The use of the Student t statistic, however, involves the assumption that the d_i values are noramlly distributed. Though it is difficult to evaluate that assumption with such a small sample, the assumption is not strongly supported by the available data. It was therefore decided to apply a nonparametric test that does not require the assumption of normality. The Wilcoxon Signed Rank Test is appropriate for this application. It is less "powerful", however, than the student t test; that is, for the same probability of a type I error (α), the probability of a type II error (β) is greater for the Wilcoxon Test than the Student t test. The null and alternative hypotheses are the same for both tests.

Wilcoxon Signed Rank Test

Let: T = the sum of the ranks of the positive or the negative differences (d_i) in paired test scores, which ever sum of ranks is smaller. Since there are no negative differences, the sum of the negative ranks = 0.

For n = 11 and T = 0, the null hypothesis is again rejected at a level of significance (α) of 0.005.

Thus, both tests found a very high level of statistical significance in the increase in test scores, lending strong support to the effectiveness of the training program.

No control groups were used in the evaluation design because of the small sample size available, and because they did not seem to be strictly required to support the contention that the increase in test scores was caused by the training. Neither a pure (no treatment) control group nor a placebo (false treatment) group was needed, since it would be extremely unlikely that historical effects, subject maturation, Hawthorne effects, etc. could have caused a significant increase in objectively measured knowledge of such technical subject material. The use of a no-pretest control group would have been helpful to test for possible bias

in post-test scores caused by pre-test sensitization. Again, however, the facts that (1) the two tests were separated by an interval of eleven weeks and (2) the pre-test was not returned to the trainees, minimizes the possibility of such bias.

In addition to its use in the evaluation process, the corrected post-tests were returned to the trainees in a follow-up meeting. The entire test was was reviewed in an attempt to clear up any remaining areas of confusion. Thus, the test was used as a teaching vehicle as well as an evaluation instrument.

SUMMARY AND RECOMMENDATIONS

The evidence presented in the previous section indicates that the training program was successful in achieving it short-range goals, all of which involved the imparting of specific knowledge and skills concerning evaluation methodologies.

The long-range value of the training program, however, can be evaluated only in terms of improvements in the quality of the evaluations conducted for programs administered under Title XX. Such improvements will require some significant administrative action in addition to technical knowledge. In that regard, the following recommendations are offered concerning program evaluation policies and procedures within Title XX Administration:

1. That the importance of <u>impact</u> evaluations (in addition to process evaluations and administrative monitoring) be impressed upon all social service program and training representatives; further that these representatives be held responsible for the technical adequacy of the evaluation methodologies employed in programs that they monitor.
2. That the Management, Planning, and Development Section of Title XX Administration in cooperation with the Contract Services Section publish:
   (a) A clear statement of intent or purpose concerning program evaluations (i.e., <u>why</u> they are to be done).
   (b) A fairly detailed set of technical guidelines as to the necessary components and characteristics of an impact evaluation (i.e., <u>how</u> they are to be done).
3. That one or more qualified individuals within Title XX Administration be identified to contractors as available sources of technical guidance in the contractor's process of:

(a) Designing an appropriate evaluation methodology for his or her program

(b) Describing that methodology in the contract proposal.

(c) Implementing the evaluation methodology during the performance of the contract.

(d) Analyzing and reporting the results of the evaluation.

APPENDIX A

Pre-Test

COMPREHENSIVE EXAM
PROGRAM EVALUATION METHODOLOGIES

Name: _____

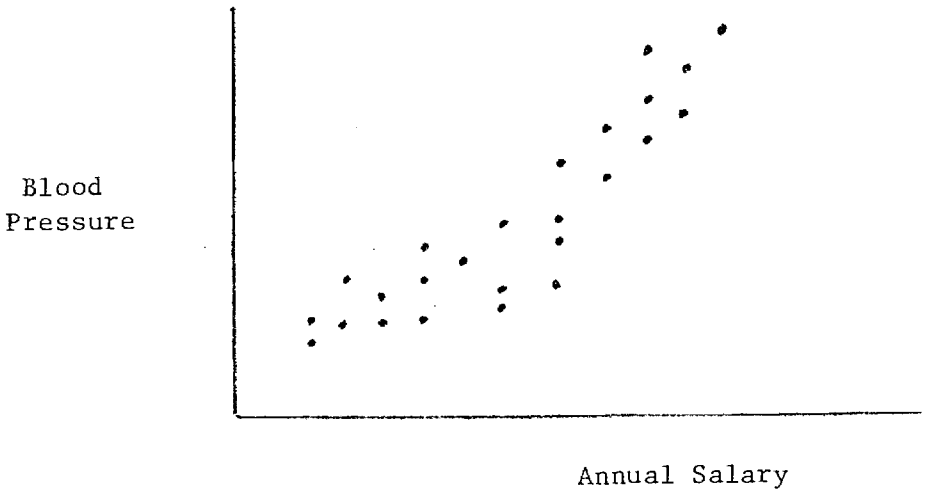
3 Pts. 1. Briefly distinguish between "process evaluation" and "impact evaluation."

3 Pts. 2. Briefly distinguish between "effectiveness" and "efficiency" as criteria for program evaluation.

3 Pts. 3. Briefly distinguish between "evaluation" and "monitoring."

4 Pts. 4. Briefly explain the meanings of "internal validity" and "external validity" of an evaluation study.

5 Pts. 5. The data shown in the graph below resulted from a study which is part of a large research program to determine the causes of high blood pressure. In this particular study, 25 managers in a single company were randomly selected. Each manager's blood pressure was measured, and his current annual salary was recorded. In the graph, blood pressure is plotted versus salary for each manager. The managers with the highest salaries occupy key positions of responsibility in this very dynamic company. What conclusions would you draw from these results and what recommendations would you make to the general public or to the researchers?



4 Pts. 6. Explain the procedure and the advantages of a "Solomon 4-group experimental design."

4 Pts. 7. Explain the "Hawthorne effect" and an experimental design that allows that effect to be meashred.

4 Pts. 8. Explain how and why a "time series design" could be used.

4 Pts. 9. Explain the problem of "regression effect" in evaluation research.

3 Pts. 10. Briefly distinguish between "direct" versus "indirect" measurements.

11. For each of the following measurements, indicate whether an interval, nominal, ratio, or ordinal scale would be involved by writing an I, N. R. or O in the space provided:

2 Pts. ____ a. A count of the number of people enrolled in a jobs training program at a given time.

2 Pts. ____ b. A six-digit identification number assigned on a random basis to participants in a given program.

2 Pts. ____ c. An item on an attitude survey to which you may respond by checking: "strongly agree," "agree," "no opinion," "disagree," "strongly disagree."

2 Pts. ____ d. The body weight in kilograms of individuals participating in a nutrition program.

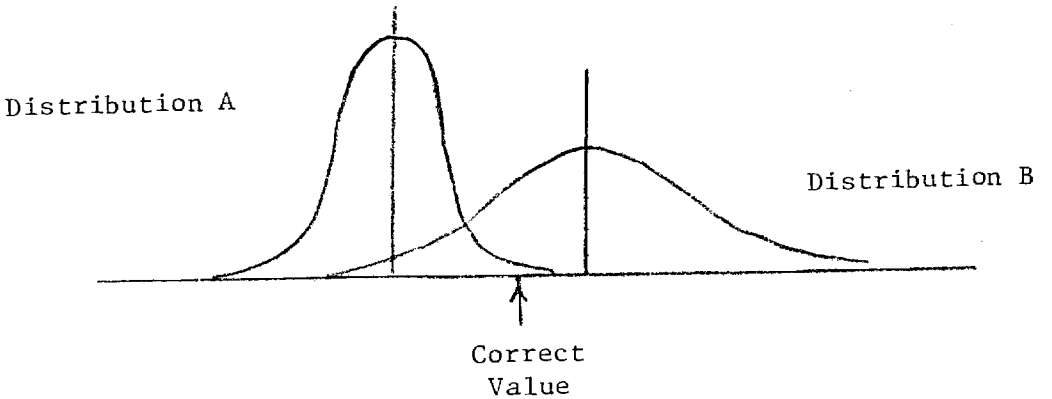
2 Pts. ____ e. Numbers assigned to children in a day care center based upon their relative height, where #1 is the ta-lest child, #2 is the second tallest child, etc.

4 Pts. 12. Briefly explain the distinguishing characteristics of "interval scales" and give one example of such a scale.

13. Two distributions of measurements of the same item are shown below along with the correct value for the measurement.

2 Pts. a. Which distribution is more precise? _____

2 Pts. b. Which distribution is more accurate? _____



4 Pts. 14. Contrast the nature of measurement errors that cause problems of inaccuracy and the nature of errors that cause imprecision.

4 Pts. 15. Explain "reliability" and "Validity" as characteristics of survey measurements.

4 Pts. 16. Explain the problem of "nonrespondent bias" in mailed questionnaire surveys.

5 Pts. 17. Suppose we wanted to determine the proportion of cigarette smokers in the United States who would stop smoking if the price of cigarettes were raised by 50%. We have taken an initial random sample of 75 smokers, 25 of whom said that they would quit smoking if such a price increase occurred. How large a random sample would we have to take in order to be 95% certain that the proportion we obtained would be within $\pm$ 2% of the proportion we would obtain if we asked all cigarette smokers in the U.S.? A table of standard normal Z values is provided below. Show your calculation and label your answer clearly.

Z	% of Values in Normal Distributions Contained Within $\mu \pm Z \cdot \sigma$
0.67	50.0%
1.00	68.3
1.65	90.0
1.96	95.0
2.00	95.5
2.58	99.0
3.00	99.7

4 Pts. 18. Explain the nature of "Type I (or alpha) errors" and "Type II (or beta) errors" in statistical tests of hypotheses.

4 Pts. 19. Briefly contrast the nature of circumstances in which "analysis of variance" is an appropriate statistical tool versus circumstances in which "correlational analysis" is more appropriate.

4 Pts. 20. Briefly describe the nature of "nonparametric" statistical procedures and the circumstances in which they are useful.

3 Pts. 21. What is SPSS?

22. Construct a CPM precedence diagram for the project described below and perform all forward, reverse, and total slack computations. Show your work clearly. Then answer the questions that follow the description of the project.

Activity A is the first activity in the project. When activity A is finished, activities B, C, and D can all begin simultaneously. Activities B and D must be completed before E can start.

Activity F can begin when C is finished, but activity G cannot start until both C and D are completed. Activity H, which is the final activity in the project, can begin when E, F, and G are all finished. The activity durations are as follows:

Activity	Duration
A	5 days
B	4
C	6
D	7
E	8
F	7
G	7
H	3

4 Pts. Network and Calculations:

2 Pts. a. What is the total duration of the project? _____

2 Pts. b. What sequence of activities comprises the critical path?

_____

3 Pts. c. How many days of total slack are associated with:

Activity B _____

Activity D _____

Activity G _____

2 Pts. d. If you needed to shorten the project duration by one day, which of the following activities would you consider shortening? (circle all that are appropriate):

A

B

E

G

APPENDIX B

Detailed Course Outline

PROGRAM EVALUATION METHODOLOGIES
Course Outline

I. Philosophy & overview of program evaluation
   - A. The program as a system of resources, activities, and results
   - B. Process evaluations vs. impact evaluations
   - C. Why evaluate programs?
     1. Management decision making
     2. Knowledge building
   - D. Criteria for program impact evaluations
     1. Effectiveness
     2. Efficiency
     3. Secondary criteria and side effects
   - E. Evaluation vs. monitoring
   - F. Characteristics of useful program objectives and goals
     1. Relevance
     2. Measurability
     3. Quantified goals for definite time frames
     4. Realistically attainable
     5. Reasonably compatible
   - G. Continuous vs. one-shot evaluations
   - H. The burden of proof - alternative assumptions
   - I. Barriers to the evaluation process
     1. Conceptual/philosophical barriers
     2. Technical/methodological barriers
     3. Organizational/political/behavioral barriers
   - J. Symptoms of problems in evaluations
   - K. Why do we fear program evaluations?
   - L. Suggestions for obtaining cooperation in the evaluation effort
   - M. Where should the evaluator be placed organizationally?
   - N. The cost of evaluation

II. Experimental design for evaluation
   - A. Internal and external validity
   - B. Limitations of correlational studies
   - C. Experimental designs
     1. Proving that change occurred - pre-treatment and post-treatment measurement
     2. Controlling for history and subject maturation
     3. Controlling for effects of pretesting and test/treatment interaction - Solomon 4-group design
     4. Controlling for Hawthorne effect - placebo treatments
     5. Time series designs with or without control groups
   - D. Other problems
     1. Regression effects
     2. Non-equivalent groups
     3. Subject mortality
     4. Instrument decay
     5. Dangers of interpolating and extrapolating relationships
     6. Dangers of generalizing results to other populations and situations

III. Measurement processes, scales, and errors
- A. Generalized definition of measurement
- B. Methods of measurement
  1. Direct vs. indirect
  2. Objective vs. subjective
- C. Measurement scale
  1. Nominal scales
  2. Ordinal scales
  3. Interval scales
  4. Ratio scales
- D. Qualities of measurement
  1. Precision
  2. Accuracy
- E. Measurement errors
  1. Random errors
  2. Systematic errors or biases
- F. Sources of measurement errors
  1. Conditions of measurement
  2. Measurement instrument
  3. Individual performing the measurement
  4. Subject of the measurement
  5. Sampling errors

IV. Questionnaire and interview surveys
- A. Qualities of survey measurements
  1. Reliability
  2. Validity
- B. Sources of survey measurement errors
- C. Conceptualization of variables
- D. Item construction: options and guidelines
  1. Item types
  2. Dangers to be avoided
  3. Ideal characteristics
- E. Formatting considerations
- F. Questionnaire mailing considerations
- G. Guidelines for interviewing
- H. Pilot tests for questionnaires and interviews
  1. General tests
  2. Testing for reliability
  3. Testing for validity

V. Attitude measurement
- A. Definition of attitude
- B. Dimensions of attitudes
  1. Cognitive
  2. Affective
  3. Behavioral
- C. Attitude measurement procedures
  1. Thurstone scale
  2. Likert scale (summated ratings)
  3. Osgood scale (semantic differential)
- D. Measuring attitudes toward jobs

VI. Sampling
- A. Selected terms and definitions
  - 1. Element
  - 2. Population
  - 3. Survey frame
  - 4. Sample
- B. Representativeness - the primary concern in sampling
  - 1. Elimination of sampling biases
  - 2. Adequate sample size
- C. Sampling designs
  - 1. Simple random sampling
  - 2. Systematic sampling
  - 3. Stratefied sampling
  - 4. Multi-stage cluster sampling
  - 5. Probability proportional to size (PPS) sampling
- D. Nonrespondent bias
- E. Statistical concepts related to sampling
  - 1. Mean
  - 2. Standard deviation
  - 3. Standard error of the mean
- F. Sample size determination
  - 1. Estimating a population mean
  - 2. Estimating a population proportion

VII. Statistical analysis - an overview
- A. General nature of statistics
- B. Hypothesis testing
  - 1. Null and alternative hypotheses
  - 2. Process of testing the hypothesis
  - 3. Type I and Type II errors
    - a. Level of significance
    - b. Power of test
- C. Analysis of variance (ANOVA)
  - 1. Applicable conditions
  - 2. Conceptual model and hypotheses
  - 3. Example
- D. Correlation analysis
  - 1. Applicable conditions
  - 2. Conceptual model and hypotheses
  - 3. Example
- E. Nonparametric and distribution-free procedures
  - 1. General characteristics as compared with parametric procedures
  - 2. Some specific procedures with examples
    - a. Contingency tables
    - b. Mann-Whitney U test
    - c. Kruskal-Wallis analysis of variance by ranks
    - d. Rank correlation

VIII. Computer support for evaluation
- A. When the computer is needed

- B. Requirements in using the computer
- C. Commercially available software
- D. Developing special purpose software
- E. Input/output considerations
- F. Data storage and backup
- G. Limitations and potential problems

IX. Managing the evaluation process
- A. Desirable qualifications and characteristics of the evaluation manager
- B. Typical task network for evaluation process
- C. Project management with CPM (Critical Path Method)
  1. Advantages of using CPM
  2. Step-by-step procedure and calculations
  3. Additional principles and concepts
     a. Updating the project plan
     b. Concept of relative criticality
     c. Project compression
     d. Leveling resource profiles

X. Overall review

APPENDIX C

Handout Materials

SELECTED RELEVANT BOOKS

1. Abert, James G. and Murray Kamrass, Social Experiments and Social Program Evaluation, Ballinger Publishing Co., Cambridge, Mass., 1974.

2. American Institute of Research, Evaluative Research: Strategies and Methods, Pittsburgh, Pa., 1970.

3. Babbie, Earl R., Survey Research Methods, Wadsworth Publishing Co., Belmont, Cal., 1973.

4. Bernstein, Ilene and Howard E. Freeman, Academic and Entrepreneurial Research: The Consequences of Diversity in Federal Evaluation Studies, Russell Sage Foundation, New York, 1975.

5. Blalock, Hubert M., Jr., Social Statistics, McGraw Hill, New York, 1960.

6. Burton, T. L. and G. E. Cherry, Social Research Techniques for Planners, Allen and Unwin, London, 1970.

7. Campbell, Donald T. and Julian C. Stanley, Experimental and Quasi-Experimental Designs for Research, Rand McNally, Chicago, 1966.

8. Caro, Francis G., Readings in Evaluation Research, Russell Sage Foundation, New York, 1971.

9. Davis, James A., Elementary Survey Analysis, Prentice-Hall, Englewood Cliffs, N.J., 1971.

10. Franklin, Jack L. and Jean H. Thrasher, An Introduction to Program Evaluation, John Wiley & Sons, New York, 1976.

11. Guilford, J., Psychometric Methods, McGraw Hill, New York, 1954.

12. Hatry, Harry P., R. E. Winnie, and D. M. Fish, Practical Program Evaluation for State and Local Government Officials, The Urban Institute, Washington, D.C., 1973.

13. Hinrichs, H. H. and G. M. Taylor, Systematic Analysis: A Primer on Benefit-Cost Analysis and Program Evaluation, Goodyear Publishing Co., Pacific Palisades, Cal., 1972.

14. Institute on Rehabilitation Services, Program Evaluation: A Beginning Statement, U.S. Department of Health, Education, and Welfare, Rehabilitation Services Administration, Washington, D.C., 1972.

15. Miller, D., Handbook of Research Design and Social Measurement, David McKay, New York, 1964.

16. National Advisory Council on Education Professions Development, Search for Success: Toward Policy on Educational Evaluation, Washington, D.C., June, 1974.

17. Oppenheim, A. N., Questionnaire Design and Attitude Measurement, Basic Books, New York, 1966.

18. O'Toole, R. (editor), The Organization, Management, and Tactics of Social Research, Schenkman Publishing Co., Cambridge, Mass., 1971.

19. Rossi, Peter H. and Walter Williams (editors), Evaluating Social Programs, Seminar Press, New York, 1962.

20. Schulberg, Herbert, Alan Sheldon, and Frank Baker, Program Evaluation in the Health Fields, Behavioral Publications, New York, 1969.

21. Sjoberg, Gideon and Roger Nett, A Methodology for Social Research, Harper & Row, New York, 1968.

22. Suchman, Edward, Evaluative Research, Russell Sage Foundation, New York, 1967.

23. Thompson, Mark S., Evaluation for Decision in Social Programmes, Lexington Books, Lexington, Mass., 1975.

24. Tripodi, Tony, Phillip Fellin, and Irwin Epstein, Social Program Evaluation: Guidelines for Health, Education, and Welfare Administrators, F. E. Peacock Publishers, Itasca, Ill., 1971.

25. Weiss, Carol H., Evaluating Action Programs: Readings in Social Action and Education, Allyn & Bacon, Boston, 1972.

26. Weiss, Carol H., Evaluation Research, Prentice-Hall, Englewood Cliffs, N.J., 1972.

27. Wholey, Joseph S., et. al., Federal Evaluation Policy, The Urban Institute, Washington, D.C., 1970.

PHILOSOPHY AND OVERVIEW OF PROGRAM EVALUATION

The Program as a System

Types of Evaluations

1. Process & Resource Evaluation

2. Results, Outcomes, or Impact Evaluation

   Which is "right"?

Why Evaluate Programs

1. We are required to do so, and funds are allocated for that purpose.
2. Decision making
   a. Within the program.
   b. Above the program.
3. Knowledge building

SOME PERTINENT QUESTIONS:

--Do you care whether your programs work?

--Is it politically expedient to admit during the process of selling a program that you are not sure whether it will work?

--Is it politically expedient to provide information on how well a program is working even when it is not meeting expectations?

Criteria for Program Impact Evaluations

1. Effectiveness

2. Efficiency or Cost Effectiveness

3. Secondary criteria:

   a. Side effects

   b. Staff development and satisfaction

   etc.

Evaluation vs. Monitoring

Characteristics of Useful Program Objectives & Goals

WHAT CAN DO MORE HARM TO THE PROSPECTS FOR FUTURE FUNDING OF SOCIAL SERVICE PROGRAMS THAN:

--Vague or non-existent objectives
--Unrealistic promises
--Unsubstantiated performance

Continuous vs. Late, One-Shot Evaluations

Do We Really Need to Specify Definite Objectives?

The Burden of Proof - Alternative Presumptions

Barriers to the Evaluation Process

Symptoms of Problems in Evaluations
(The causes may not be as obvious as they seem.)

Why Do We Fear Program Evaluations?

--Would you rather not have your program evaluated?
--Can we separate program evaluation and personal performance evaluation?

Suggestions for Obtaining Organization Cooperation for the Program Evaluation Effort

Where Should the "Evaluator" Be Placed Organizationally?

1. Outside consultant or researcher.

2. Internal staff reporting to a level higher than program administrator.

3. Internal staff reporting to program administrator, but uninvolved in program operation.

4. Internal staff also involved in program operation.

How Much Should Evaluation Cost?

MINI-CASES FOR SMALL GROUP WORK SESSION

For each of the hypothetical, federally subsidized programs described below:

1. Identify several (at least two or three) criteria that would be relevant to an impact evaluation of the program.
2. Specify how each of those criteria could be measured, and describe any problems that you would anticipate with respect to each measurement.

Each group should select a spokesman who will be prepared to report to the class.

PROGRAMS:

A. A municipal level program to train police patrolmen in techniques of intervention in family disputes.

B. A state level educational program for juvenile offenders that allows participants to earn early releases from detention centers by satisfactorally completing certain academic and vocational course work.

C. A community level child care program for welfare families that operates on the following weekday schedule:
   Ages 6 months - 5 years: 8:00 AM - 5:30 PM.
   Ages 6 years - 12 years: 2:00 PM - 5:30 PM.
   The program includes the following services:
   1) Transportation to the center from home or school, and transportation back home in the evening.
   2) A nutrition program that provides breakfast, lunch, and an afternoon snack for the younger children, and the afternoon snack for the older children.
   3) Supervised indoor and outdoor play. Educational resources available for all ages and limited planned educational activities for the younger children.
   4) Monthly visits by a doctor to monitor the general health of each child and to administer innoculations with prior parental consent.

D. A community level program for senior citizens. The only criterion for eligibility is that the individual be at least 65 years old. Participants pay a monthly fee ranging from $1 to $20, depending upon their financial status. Services include:
   1) A Senior Citizen's Activity Center that has facilities for games, reading, watching TV, working on craft projects, or simply relaxing with friends. Center is staffed by a salaried director and one other paid employee, who are supplemented by volunteer, unpaid program participants.
   2) Regularly scheduled transportation between the Activity Center and participants homes; also between the Center and a nearby shopping center and medical complex.

3) Occasional special outings to entertainment events, etc. Transportation and admission provided through the program.
4) Meetings every Wednesday evening to plan future activities. Programs involving guest speakers are included for about half the meetings.
5) Other organized activities as developed by the program participants, including charitable and community service endeavors.

EXPERIMENTAL DESIGN FOR EVALUATION

Validity of Evaluation Studies

1. Internal Validity

2. External Validity

Limitations of Correctional Studies

Example:

Three Ways to Interpret Results:

Experimental Designs

1. Proving that change occurred (pre-treatment and post-treatment measurement.

2. Controlling for "history" and "maturation." (pure control groups).

3. Controlling for effects of pre-treatment measurement and measurement/treatment interaction. (Solomon 4-group design).

4. Controlling for "Hawthorne effect." (placebo treatment).

5. Time series designs with or without control groups. (a good hybrid design.)

Other Problems

1. Regression effects

2. Non-equivalent groups

3. Subject "mortality"

4. Instrument decay

5. Dangers of interpolating & extrapolating results

6. Dangers of generalizing results to other populations or situations.

MEASUREMENT PROCESSES, SCALES AND ERRORS

Measurement: A process of assigning "symbols" to "objects" according to rules, so that the symbols represent characteristics of the objects.

Subjects of Measurement

Methods of Measurement

Scales of Measurement

1. Nominal Scales

2. Ordinal Scales

3. Interval Scales

4. Ratio Scales

Qualities of Measurement (Accuracy and Precision)

Types of Measurement Errors

1. Systematic errors or biases

2. Random errors

Sources of Measurement Errors

1. Conditions of Measurement

2. Measurement instrument

3. Individual performing the measurement

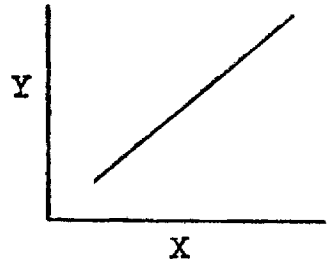
4. Subject of measurement

QUIZ - 11/23/76

Name:

1. Explain the meanings of "internal validity" and "external validity" of a research study.

2. Give three possible explanations for the cause of the relationship shown between variables X and Y below.



3. Diagram and explain the usefulness of the Solomon 4-group research design.

4. Briefly explain the problem of "regression effect" in evaluation research.

5. For each of the following measurements, indicate whether an interval, ordinal, ratio, or nominal scale would be involved by writing an I, O, R or N in the space provided:

   ___ a. The age of an individual in years.

   ___ b. Evaluation of an individual's job performance as "outstanding, good, fair, or poor".

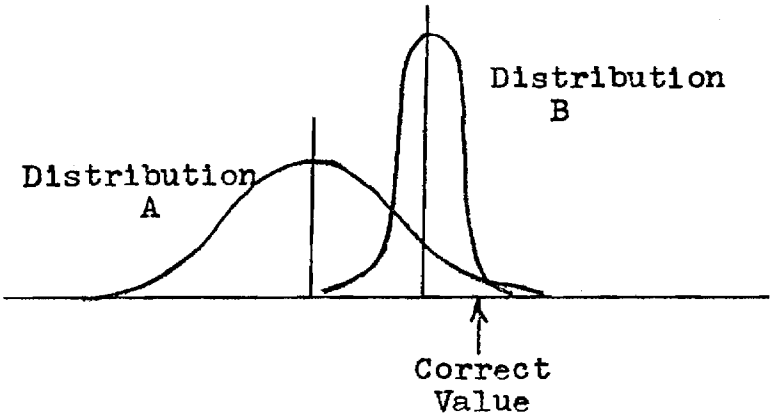
   ___ c. A question requiring a "yes" or "no" response.

   ___ d. A count of the number of people in a group who answer "yes" to a given question.

6. Given the two distributions of measurements shown below:

   a. Which distribution is more accurate? ____

   b. Which distribution is more precise? _____



QUESTIONNAIRE & INTERVIEW SURVEYS

Typical Characteristics of Survey Measurements

Qualities of Survey Measurements

1. Reliability

2. Validity

Sources of Survey Measurement Errors

Conceptualization of Variables

-In many cases, the variable is defined by the measurement.

-In many cases, the researcher cannot "collect" data; he must "create" data.

Item Construction: Options & Guidelines

1. Item types

   a. Statements vs. questions

   b. Open-ended vs. closed-ended questions

   c. Multiple choice items

   d. Ranking items

2. Avoid:

   a. Unclear or ambiguous terms and phrases.

   b. Negative phrasing.

   c. "Double-barreled" questions.

   d. Biased, leading, or unnecessarily emotion-charged terms, phrases, and associations.

3. Items should be:

   a. As short as possible, consistent with clarity.

   b. Relevant to respondents' interests, activities, concerns.

   c. Within the respondents' competence to answer.

Formatting Considerations

1. Quality of type-setting and reproduction; liberal "white space"

2. Introduction; general and specific instructions.

3. Contingency flows.

4. Ordering of activities.

Questionnaire Mailing Considerations

1. Sending: first class vs. bulk mail.

2. Returning: pre-stamped vs. business reply.

3. Follow-up Mailing.

Guidelines for Interviewing

1. Interviewer is a "neutral medium".

2. Appearance & demeanor of interviewer.

3. Familiarity with instrument, procedures, etc.

4. Follow script wording as exactly as possible.

5. Record responses exactly; make supplemental observations.

6. Probe for responses without raising tension or biasing.

7. Supervision: regular reporting & editting.

8. Reassignment of interview subjects.

9. Verification of interviews.

10. Termination of interviewers.

Pilot Tests of Questionnaires & Interviews

1. General tests.

2. Testing for reliability.

3. Testing for validity.

ATTITUDES & ATTITUDE MEASUREMENT

Definition

A learned predisposition to respond to a given object in a favorable or unfavorable way.

Dimensions of Attitudes

1. Cognitive

2. Behavioral

3. Affective

Attitude Measurement Procedures

1. Thurstone Scale

2. Likert Scale (Summated Ratings)

3. Osgood Scale (Semantic Differential)

Measuring Attitudes Toward Jobs

Factors to Consider:

Relative Importance of Factors:

SAMPLING

Selected Terms & Definitions

1. Element-

2. Population-

3. Survey frame-

4. Sample-

The Primary Concern in Sampling

"Representativeness"

Reduction of "sampling error" through:

1. Elimination of sampling biases.

2. Adequate sample size.

Sampling Designs

1. Simple random sampling

2. Systematic sampling

3. Stratefied sampling

(implied stratefication through systematic sampling)

4. Multistage cluster sampling (unstratefied)

(# of clusters vs. # of elements per cluster)

5. Multistage cluster sampling (stratefied)

6. Probability proportional to size (PPS) sampling

Nonrespondent Bias

QUIZ - 12/14/76

1. Explain the "Hawthorne Effect," and describe a research design to measure that effect.

2. Distinguish between "reliability" and validity" as qualities of survey measurements.

3. (a) Identify and briefly describe the three "dimensions" of attitudes that we have discussed.

   (b) Identify (do not describe) three well-known types of attitude measurement scales.

4. Briefly explain each of the following sampling procedures:
   (a) Systematic sampling

   (b) Stratefied sampling

   (c) Multistage cluster sampling (unstratefied)

5. Explain the problem of "nonrespondent bias."

STATISTICAL CONCEPTS RELATED TO MEASUREMENT

Mean

Given a series of measurements of the same object yielding the values

$$x_1, x_2, x_3, \ldots, x_n,$$

The MEAN of those values is computed as:

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

Standard Deviation

The STANDARD DEVIATION of the distribution of values (which is a measure of PRECISION) is computed as:

$$\sigma_s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{n-1}}$$

(Note: The symbol "σ_s" is used here to be consistent with the readings. The more popular symbol is "s".)

Assuming that the measurements are randomly and NORMALLY distributed, then for any given value of Z, we can determine the percentage of the measurement values that fall within the range:

$$\bar{x} - Z \cdot \sigma_s \text{ to } \bar{x} + Z \cdot \sigma_s$$

Similarly, given a particular percentage of measurement values, we can determine the number (Z) of standard deviations (σ_s) on either side of the mean ($\bar{x}$) that will enclose that percentage of values. Z values and associated probabilities are given in Standard Normal Tables. Examples are shown on the next page.

Z	% of Values Contained within $\bar{x} \pm Z \cdot \sigma_s$
0.67	50%
1.00	68.3
1.65	90
2.00	95.5
3.00	99.7

Standard Error of the Mean

If several samples of measurements, each containing n measurements, were taken and $\bar{x}$ were computed for each sample, there would be some variation in the values for $\bar{x}$ from sample to sample. The values for $\bar{x}$ would be normally distributed and would have a STANDARD ERROR (σ_m) which can be estimated from a single sample of n measurements by the computation:

$$\sigma_m = \frac{\sigma_s}{\sqrt{n}}$$

The standard error of the mean (σ_m) is conceptually the standard deviation of a distibution of sample means ($\bar{x}$) computed from separate samples of n measurements each.

DETERMINING REQUIRED SAMPLE SIZES

For Estimating a Population Mean (μ)

Suppose that we are interested in estimating the mean (μ) of a population. In terms of our discussion of measurement, we can think of μ as the $\bar{x}$ value that we would compute from a sample containing an infinite number of measurements of the same object. Thus, as our sample size (n) approaches infinity, the value of $\bar{x}$ would approach μ.

We wish to determine how large a sample (n) we must take so that we can have a given level of confidence (C) that the value of $\bar{x}$ computed from that sample is within some specified range (E) of μ.

$$\text{Probability } (\bar{x} \text{ is within } \mu \pm E) = C$$

Substitute $Z \cdot \sigma_m$ for E to represent the range of allowable error. For the given level of confidence (C), we can determine a corresponding Z value from the Standard Normal Table such that

$$\text{Probability } (\bar{x} \text{ is within } \mu \pm Z \cdot \sigma_m) = C$$

We know that $\sigma_m = \frac{\sigma_s}{\sqrt{n}}$, so

$$E = Z \frac{\sigma_s}{\sqrt{n}}$$

Given that we have specified the values of E and Z, and we have an estimate of σ_s, we can solve for the necessary sample size as:

$$n = \frac{Z^2 \sigma_s^2}{E^2}, \quad \begin{array}{l}\text{(round up to next} \\ \text{higher integer)}\end{array}$$

For Estimating a Population Proportion (π)

Suppose that we are interested in estimating the proportion (π) of items in a population that fall within a given class. By taking a sample of the population we can obtain an estimate (p) of the proportion. As

the sample size (n) increases the value of p would approach π.

We wish to determine how large a sample (n) we must take so that we can have a given level of confidence (C) that the value of p computed from the sample is within some specified range (E) of π.

$$\text{Probability (p is within } \pi \pm E) = C$$

The standard error (σ_p) of the sample proportion (p) is estimated as:

$$\sigma_p = \sqrt{\frac{(p)\ (1-p)}{n}},$$

where p is the sample poportion obtained from a sample of size n.

Since the sample proportions (p) are approximately nornally distrubuted for large sample sizes, we can substitute $Z \cdot \sigma_p$ for E and find the Z value in the Standard Normal TAble such that

$$\text{Probability (p is within } \pi \pm Z \cdot \sigma_p) = C$$

Thus,

$$E = Z \cdot \sqrt{\frac{(p)\ (1-p)}{n}}$$

Given that we have specified the values of E and Z, and we have an estimate of p, we can solve for the necessary sample size as:

$$n = \frac{Z^2 (p)\ (1-p)}{E^2},$$

(Express p and E as decimal fractions & round answer up to next higher integer)

If we have no estimate of p, use p = 0.50 to give the larget possible value of n.

Homework Problems

1. Suppose that we are attempting to measure the speed of Jimmy Conners' serve. We have taken 30 measurements and have obtained the following values in miles per hour.

111.3	113.2	112.5	111.9	112.3
113.6	112.5	111.6	113.2	111.6
112.4	117.6	113.1	112.0	110.0
107.2	110.2	110.7	115.6	113.3
106.1	116.7	114.3	107.8	109.8
110.9	112.1	108.9	116.1	114.1

   a) Determine the mean, standard deviation, and standard error of the mean.

   b) How large a sample will we need to take if we want to be 90% certain that our sample mean falls within ±0.5 miles per hour of the population mean?

2. We are interested in determining the proportion of adults (at least 21 years old) living within the State of Georgia who are not citizens of the U.S. We have taken a random sample of 100 and found 7 non-U.S. citizens. How large a sample must we take if we want to be 90% certain that our sample proportion is within ± 1% of the population proportion?

STATISTICAL PROCEDURES

The General Nature of Statistics

Hypothesis Testing

1. "Null" and "Alternative" Hypotheses

2. Process of Testing the Hypothesis

3. "Type I" and "Type II" Errors

"Level of Significance"

"Power of Test"

Analysis of Variance (ANOVA)

1. Applicable Conditions

2. Models and Hypotheses

3. Example

Correlation Analysis

1. Applicable Conditions

2. Models and Hypotheses

3. Example

Nonparametric & Distribution - Free Procedures

1. General nature of these procedures as compared with parametric procedures.

2. Some specific procedures:

   a. Contingency Table (& Median Test)

b. Mann-Whitney U Test

c. Kruskal-Wallis Analysis of Variance by Ranks

d. Rank Correlation

COMPUTER SUPPORT FOR EVALUATION

When Computer Is Needed

Requirements for Computer Usage

Available Software

Developing Special Software

Input/Output Considerations

1. Input - Minimize the labor of preparing data for computer input consistent with good measurement procedures:

2. Output - Minimize the labor of preparing, duplicating, and reading management reports;

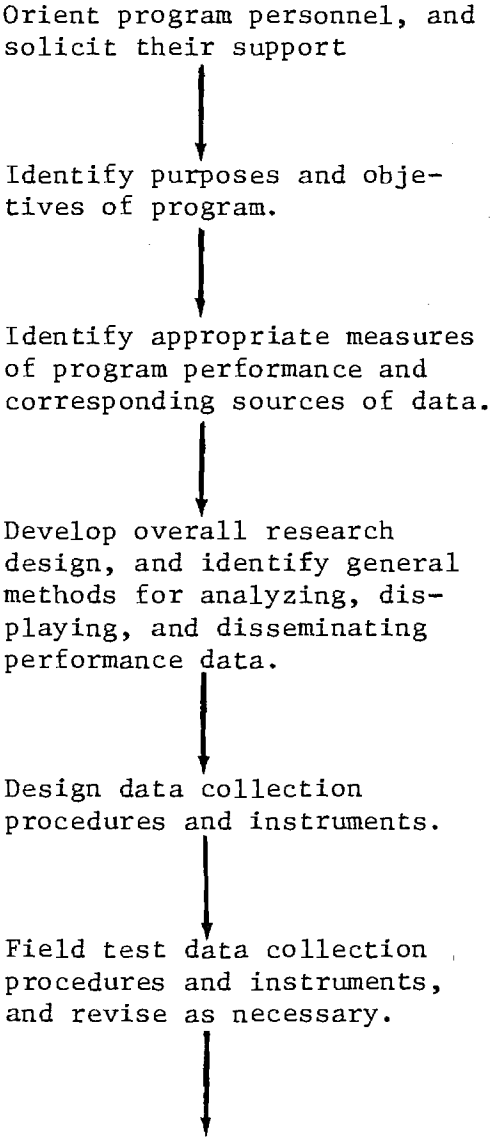
Backup Considerations

Limitations and Potential Problems

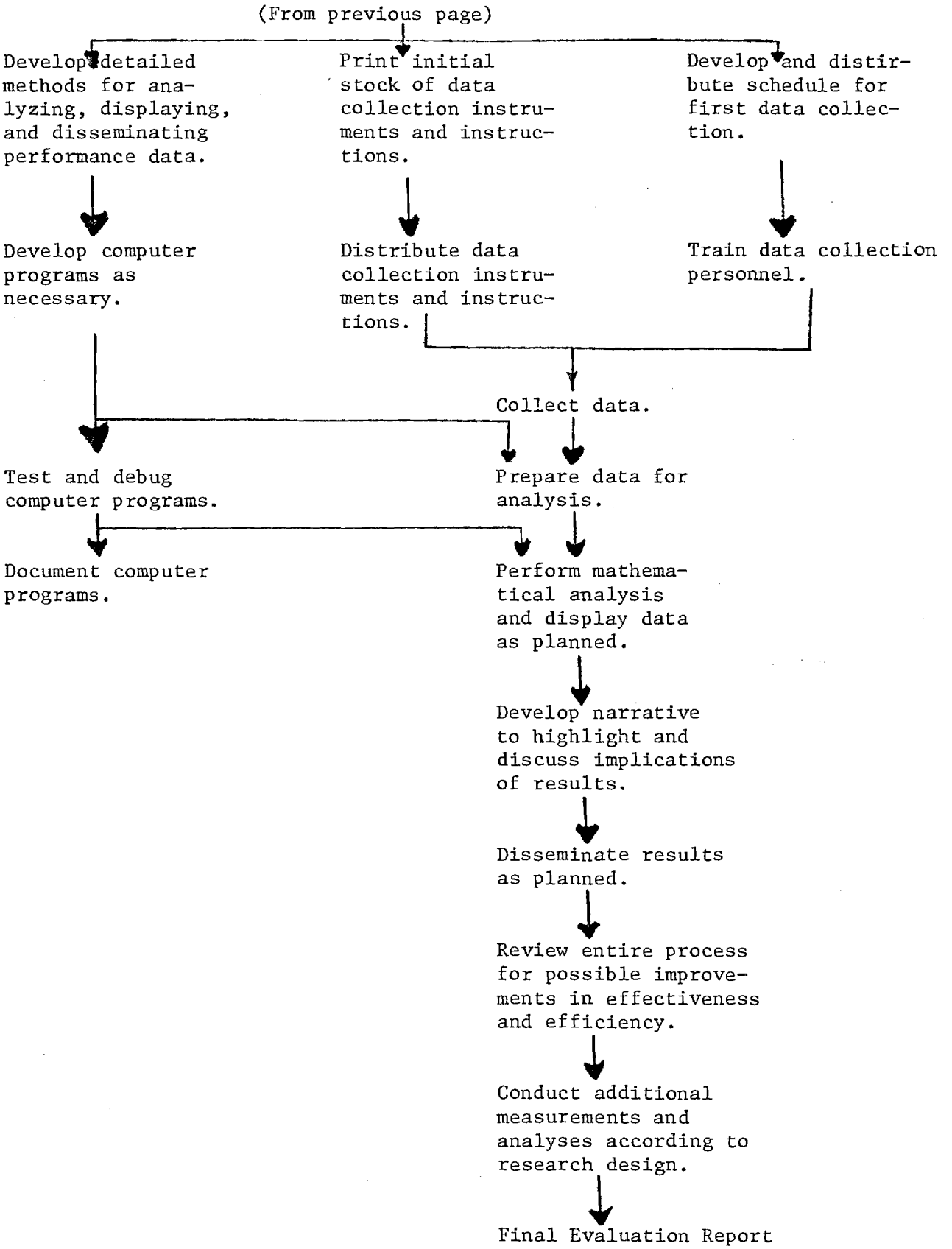
MANAGING THE EVALUATION PROCESS

Desirable Qualifications & Characteristics of Evaluation Manager

Typical Task Network For Evaluation Process



(continued next page)


(From previous page)
Develop detailed methods for analyzing, displaying, and disseminating performance data.
Print initial stock of data collection instruments and instructions.
Develop and distribute schedule for first data collection.
Develop computer programs as necessary.
Distribute data collection instruments and instructions.
Train data collection personnel.
Collect data.
Test and debug computer programs.
Prepare data for analysis.
Document computer programs.
Perform mathematical analysis and display data as planned.
Develop narrative to highlight and discuss implications of results.
Disseminate results as planned.
Review entire process for possible improvements in effectiveness and efficiency.
Conduct additional measurements and analyses according to research design.
Final Evaluation Report

PROJECT MANAGEMENT WITH CPM

CPM - Critical Path Method
PERT - Program Evaluation and Review Technique

Advantages of Using CPM

1. Forces you to plan the project in detail.

2. Uncovers problems in advance and helps in developing solutions.

3. Improves project-related communications by providing a simple and precise language.

4. Provides a basis for effective control.

Basic Procedure

1. Specify project objectives and constraints

   Considerations
   a. Time
   b. Cost & other resorces (e.g., personnel)
   c. Quality

2. Identify necessary activities and assign responsibility for each.

   Activity - A time consuming element of a project that has identifiable starting and completion points. It may or may not consume project resources (e.g., waiting for questionnaires to be returned consumes time but no resources).

   Level of Detail
   a. Depends on level of control desired.
   b. Any obvious change in:
      (1) the nature of the work,
      (2) the type of resource used, or
      (3) the individual responsible should signal the start of a new activity.
   c. The durations of most activities should be in the range of 1% - 20% of the total expected project duration.
   d. When in doubt, use greater detail.

3. Analyze precedence relationships among activities.

Types of Precedence Relationships

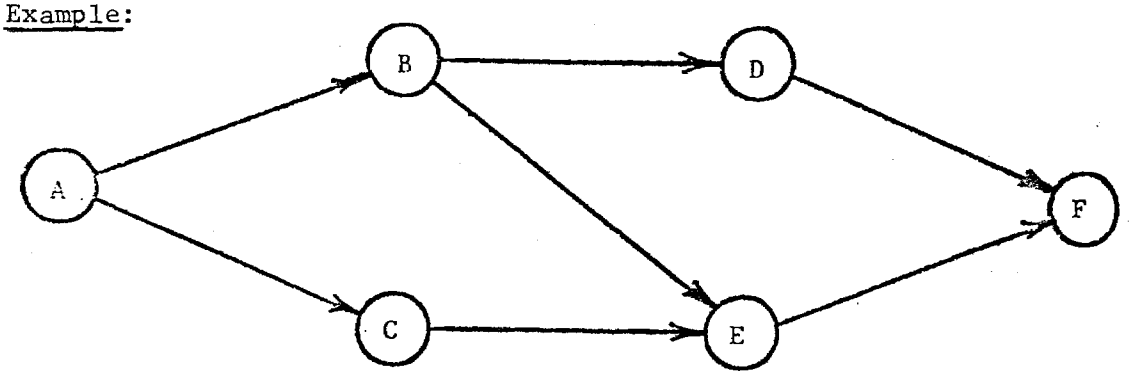
a. Technical - It is technically impossible for activity B to begin before activity A is complete.
b. Policy or Preference - It will be more efficient, safer, or will yield better quality if activity B is not started until activity A is complete.
c. Restricted Resources - Activity B cannot be started until activity A is completed, because the same people, equipment, space, etc. is required for both activities.

4. Construct project network.

Networking Schemes

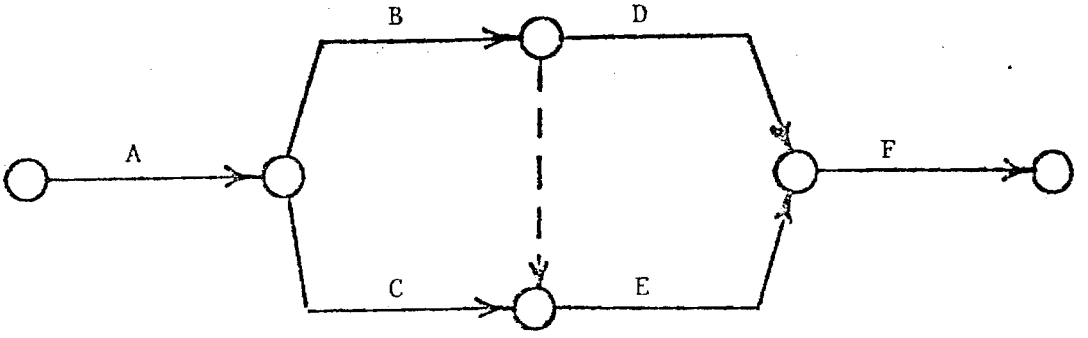
a. Activities - on - Nodes (or Precedence Diagrams): Easier but less common scheme.

Example:



b. Activities - on - Arrows: Originally developed scheme and still most common. More difficult to develop, but lends itself to plotting network on a time scale.

Equivalent example:



5. Estimate activity durations (usually in working days) based upon the most efficient (or most likely) quantity of resources assigned.

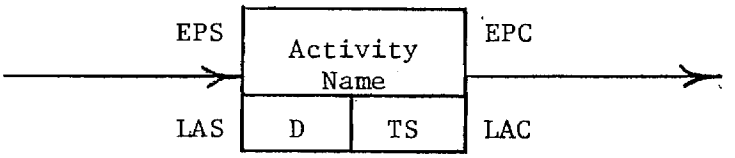
Example

No. of People Assigned	Activity Durations	Resources Consumed
1	10 work days	10 man-days
2	4 work days	8 man-days
3	3 work days	9 man-days
4	3 work days	12 man-days

On an activity-by-activity basis, include an allowance for common contingencies that could prolong the duration of each activity (e.g., computer down time could prolong a program debugging or data analysis activity; absences of personnel can delay activities).

6. Perform forward and reverse calculations.

Legend and Definitions



D - Duration of the activity
EPS - Earliest Possible Starting Time
EPC - Earliest Possible Completion Time
LAS - Latest Allowable Starting Time
LAC - Latest Allowable Completion Time
TS - Total Slack

Calculation Procedures

a. Forward calculations (EPS & EPC)
   (1) The EPS for the first activity in the network is usually set at zero.
   (2) The EPS for any other activity is the largest (or lastest) of the EPC values for all immediately preceding connected activities.
   (3) The EPC for any activity is computed as EPS + D for that activity.

b. Reverse calculations (LAS & LAC)
   (1) The LAC for the last activity in the network is usually set equal to the EPC for that activity, or it is set equal to some required completion deadline.
   (2) The LAC for any other activity is the smallest (earliest) of the LAS values for all immediately following connected activities.

(3) The LAS for any activity is computed as LAC-D for that activity.

Note: When calculations are complete, LAS-EPS for the first activity in the network should equal LAC-EPC for the last activity in the network.

7. Perform total slack computations.

Total Slack - The amount of time by which the completion of the activity can be delayed beyond its EPC without causing the project to exceed its latest allowable completion time (i.e., the LAC of the last activity in the project).

For any given activity, TS = LAC - EPC.

8. Determine the location(s) of the "critical path(s)."

Critical Path - A connected sequence of activities from the initial activity to the final activity whose total duration is equal to the total duration of the project (i.e., it is the sequence of activities that determines the duration of the project). There will always be at least one continuous critical path through the enitre project. There may be several parallel critical paths. Also, critical paths can split apart and/or merge together. Any delay in any activity on a critical path will cause the project to be delayed beyond its earliest possible completion time (i.e., the EPC of the last activity in the project).

To find the critical path(s), determine the smallest TS value in the network. All activities whose TS values are equal to that minimum value are on one or more critical paths. Determine what continuous paths involving only those activities have a total duration equal to the project duration.

Additional Principles and Concepts

1. At any given time during the performance of the project, the project plan (network) can be updated by simply treating the remaining portion of the project as if it were a new project.

2. Since the critical path(s) determine the total duration of the project, the activities on those paths should be managed very closely. However, delays in the performance of other activities can cause other paths to be lengthened to the extent that they may become more critical (i.e., longer) than the original critical paths.

3. If you desire to shorten the duration of a project by shortening activities, it is only useful to shorten activities that are on critical paths. Furthermore, it is necessary to shorten all the critical paths (or paths that become critical) in order to shorten the project.

4. It is sometimes helpful to delay the start of non-critical activities beyond the EPS (i.e., take advantage of their slack) in order to level out the quantities of resources (especially people) required over the course of the project.

EXAMPLE PROBLEM

Activity	Immediately Preceding Activities	Activity Duration
A	-	8
B	A	5
C	A	8
D	A	10
E	B	6
F	B,C,D	12
G	B,C	10
H	D	6
J	E,F,G	8
K	G,H	12

Network & Calcualtions:

APPENDIX D

Masters Thesis Proposal
for
Mr. Harry Tomas

"A Model for the Economic Evaluation
of Sliding Fee Systems
for Day Care Services"

INTRODUCTION

1.1 Background

The topic for the thesis was developed in cooperation with Mr. Lee Cash, Research Associate of the Georgia Department of Human Resources, Title XX Administration. The Department of Human Resources has been interested in developing a sliding fee system for day care centers to be implemented throughout the State of Georgia. A sliding fee system is the mechanism by which participating families share the cost of child care service with the government according to their ability to pay. The problem is that of allocating scarce resources efficiently throughout the State of Georgia to a designated population segment. These scarce resources are the limited federal funds that are allocated for day care services and the limited number of facilities and staff personnel. The recipients are families within certain family income and family composition ranges. Little is known, however, about the impact of various forms of sliding fee systems on resource usage, services provided, or costs incurred. This thesis will approach these problems and develop a quantitative model to assist the Georgia Department of Human Resources in evaluating alternative sliding fee systems.

1.2 Statement of Problem

Before 1976, a child in Georgia could obtain free day care if the gross family income was less than 61% of Georgia's median family income adjusted for family size. For example, the median income for a family of four was $1,142 per month yielding a 61% cut-off point of $697 per month. If the family gross income exceeded the above cut-off amount by one or more dollars a month, the child had to be taken from the free public day care and placed in a private center, at a cost of $100 or more per month. The abrupt cut-off of free day care service produces a dysfunctional "notch effect". The notch effect creates a disincentive for families to increase their income. It has a paradoxical effect of loss in real net income incurred over a certain income range. As income rises, the family experiences a loss in benefits in cash or in kind for which they were previously eligible. A family might lose public housing, cash assistance payments, social security supplemental income, and medicaid subsidies worth over $1,000 a year, if earned income increases from $4,300 to $4,400 (i.e. an increase of only $100). Other potential losses include Aid to Families with Dependent Children, food stamps, and increased income tax. Hence, the notch effect is the resultant drop in the family's real net income associated with the rise in earned income. With or without a sliding fee scale for day care

payments, the notch effect can take place due to these other economic losses. The sliding fee schedule, however, is designed to ameliorate the notch effect as far as the cost of day care services is concerned.

The administrative mechanism for a sliding fee system is set forth in H.R. 1, Title XX, Part B, Section 2134 (a) (2) instructing the Secretary of Health, Education and Welfare to "prescribe such schedule or schedules as may be appropriate for determining the extent to which families are required (in the light of their ability) to pay the costs of child care for which provision is made under section 2112 (a) (1)". The fee schedules can then be translated into relationships that will determine how the full cost of child care is to be shared between public funds and the participating families using child care services under the welfare reform legislation (H.R. 1).

The purpose of sliding fee scales are:

(a) To allocate scarce resources equitably and efficiently among many recipients.

(b) To encourage families to increase their earned income to the point of self-sufficiency, or at least to avoid creating disincentives for doing so..

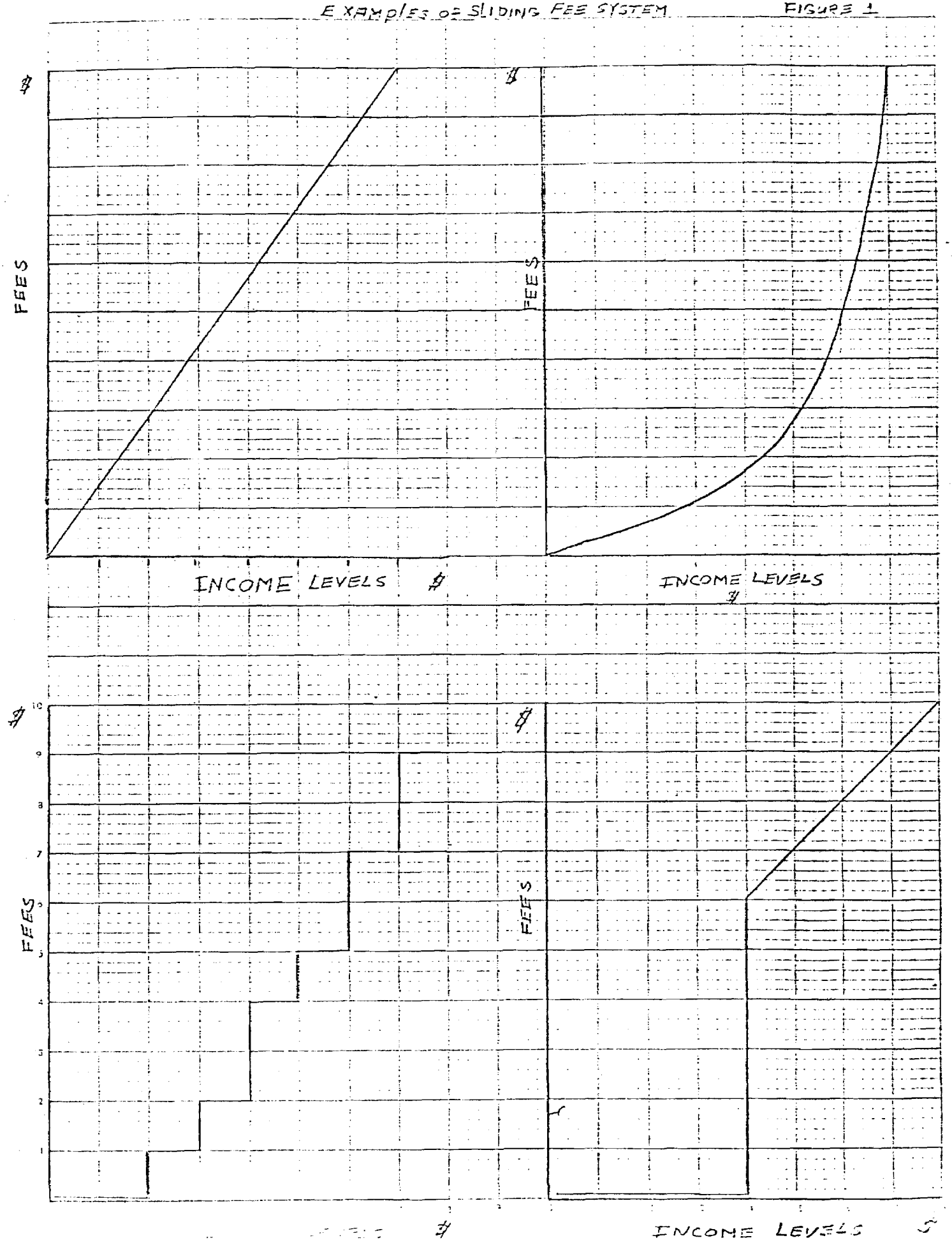
(c) To achieve equity in fee payments among those families

that will be sharing the costs of day care.

Some examples of possible sliding fee scales are graphically portrayed in Figure 1.

In January of 1977 the State of Georgia implemented an "Interim Fee System". This system is a pilot program implemented on a temporary basis until a permanent sliding fee system is designed. Under the Interim System, the State provides free day care services to approximately 11% of the 104,000 children whose family income is less than 61% of the state's median family income adjusted for family size in accordance with federal regulation. The maximum gross income for a family of four to receive completely free day care is $607 per month or $3,364 annually. The families earning between 61% and 80% (or $698 and $911 per month) are obliged to purchase the day care service in accordance with the Interim Fee System. Families eligible for the Interim Fee System are those that were participating in Title XX day care service after October 1, 1976 or new families whose income falls below the 61%-of-median level. New families earning between 61% and 80% of median income (i.e. $697 per month to $911 per month) desiring to use the public day care centers are not eligible to enter the system. Any family earning over 80% of median

EXAMPLES OF SLIDING FEE SYSTEM FIGURE 1



income does not qualify to participate in the Interim Fee System program.

The child care fee scale is depicted in Table 1-Supplement to Income Scale I. An example of the interim fee schedule for a family of four is depicted in Figure 2.

It is hoped that the present Interim Fee System will provide a smooth transition from free to full cost day care service. This Interim Fee System took into account family size and income. The 11% (12,000 children) level of service would not be reduced under the Interim System. The funds generated from this system are being used to reduce the Federal Financial Participation.

A major problem in the development of sliding fee scales is the lack of methodologies, models, and data required to evaluate the economic impacts of alternative fee scales on both the client population and the government. The problem is complex, even in the short run. To evaluate the long range impacts, it is necessary to recognize that the system exists in a dynamic environment. The number and characteristics of eligible children is constantly changing, as changes occur in family size and income distributions. The level of supply of government subsidized day care services is also likely to increase over time. The State's cost per child for providing the day care service can also be expected to inflate. The

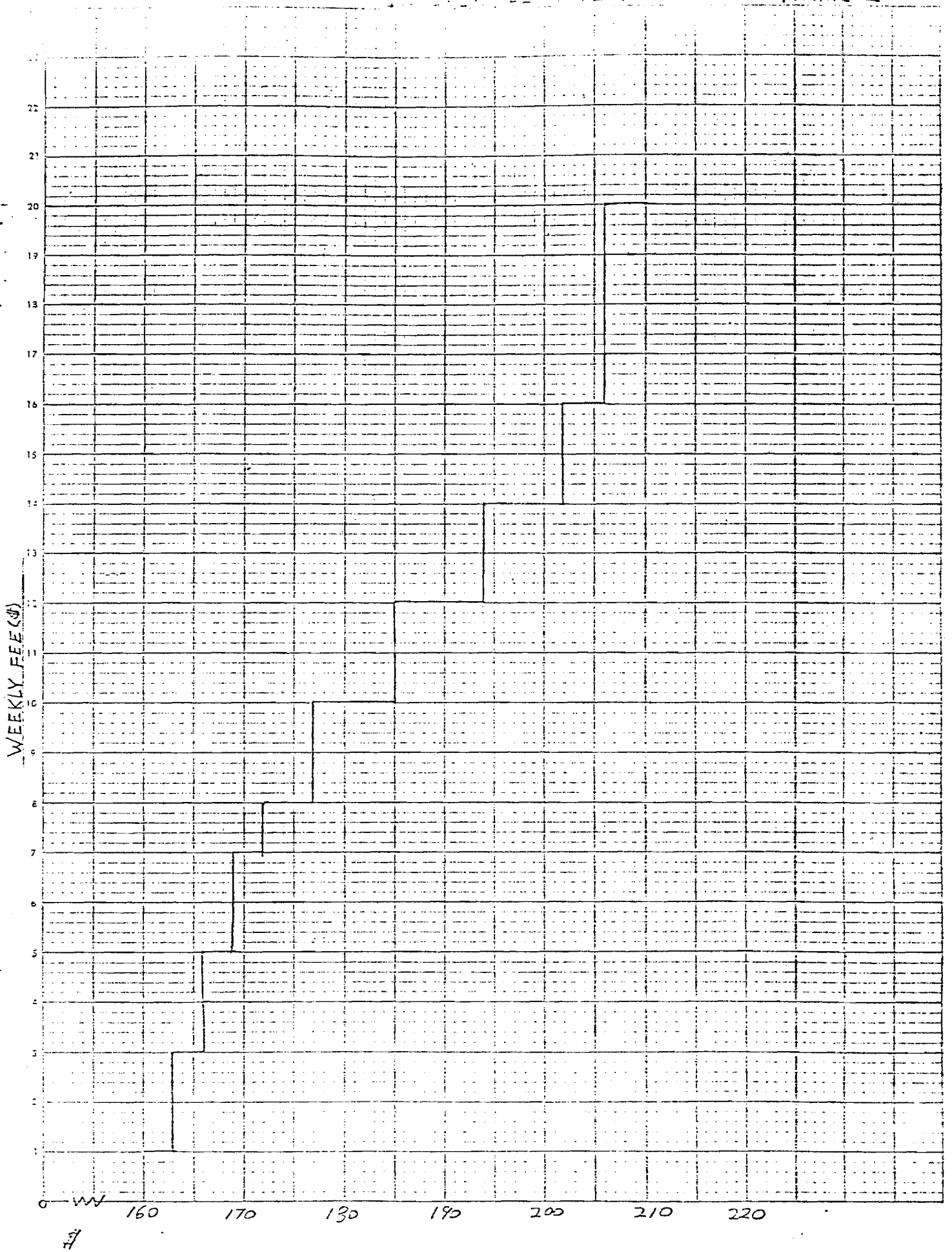
TABLE I

SUPPLEMENT TO INCOME SCALE I

CHILD CARE FEE SCALE

STEP	WEEKLY FEE	GROSS WEEKLY FAMILY INCOME FOR FAMILIES WITH									
		1 Member	2 Members	3 Members	4 Members	5 Members	6 Members	7 Members	8 Members	9 Members	10 Members
1	$ 1	83.42-84	109.09-111	134.75-137	160.41-163	186.08-189	211.75-215	216.56-220	221.37-225	226.18-230	230.99-
2	3	85-86	112-113	138-139	164-166	190-193	216-219	221-224	226-229	231-234	236-
3	5	87-88	114-115	140-142	167-169	194-196	220-223	225-228	230-233	235-238	240-
4	7	89-90	116-117	143-144	170-172	197-200	224-227	229-232	234-237	239-243	244-
5	8	91-92	118-120	145-149	173-177	201-205	228-234	233-239	238-244	244-250	24
6	10	93-96	121-126	150-155	178-185	206-215	235-244	240-250	245-255	251-261	25
7	12	97-101	127-132	156-163	186-194	216-225	245-256	251-262	256-268	262-274	26
8	14	102-105	133-137	164-170	195-202	226-234	257-267	263-273	269-279	275-285	280-
9	16	106-107	138-140	171-173	203-206	235-239	268-272	274-278	280-284	286-290	292-
10	20	108-109.03	141-142.58	174-176.12	207-209.67	240-243.22	273-276.76	279-283.06	285-289.34	291-295.63	298-301.

FIGURE 2


WEEKLY FEE ($)
0
160
170
180
190
200
210
220
GROSS WEEKLY FAMILY INCOME
FAMILIES WITH 4 MEMBERS

dynamic nature of the problem strongly suggests the applicability of a simulation approach.

1.3 Objective, Scope and Limitations

The objective of the thesis is to develop a simulation model to assist the Georgia Department of Human Resources in the economic evaluation of alternative fee scales. The simulation model will be designed as a management tool in which DHR can easily adjust the initial conditions, fee scale parameters, and other economic and demographic assumptions. The model will then provide the computational mechanism required to:

(a) Analyze the economic impact on real income of the participating families.

(b) Analyze the economic impact to the State of Georgia.

(c) Analyze the impact on the demand and supply curves of child care services.

The model is an abstraction of a real world problem which is very complex. The basis for the child care fee system is the Title XX section of the Social Security Act. Any changes in public law and policy could require changes in the basic structure of the model here being

developed. The State of Georgia has flexibility to implement their policies and concepts of child day care services. The administration of such a complex program can create regulations that could modify the model's ability to represent the real world situation. The lack of available data on child care creates the need for many assumptions and projections that would necessitate modifications as additional information becomes available. The model will be designed to facilitate user changes in the data base and model parameters.

The model will provide the framework and mechanism for DHR to evaluate several sliding fee systems. DHR must provide data they wish to use. The primary source of data that will be used in developing the model will be the 1970 Census ~~of the Bureau~~ and other information received from the Georgia Department of Human Resources.

The problem addressed in this thesis is not unique to the State of Georgia. Virtually every state is faced with the problem of designing sliding fee scales in a complex environment. Though the model developed for Georgia may not be directly applicable to any other state, the methodological concepts and the overall organization of the model should be of value to other states seeking to develop similar models.

1.4 Literature Survey

A literature survey was conducted to locate information on the methodology of sliding fee scale systems. In the general field of day care, articles were found through the use of the Science Citation Index, Social Science Index and Public Affairs Information Services Bulletin. The articles dealt with the operation of day care programs, cost effectiveness, status of day care and the need for day care services. Emphasis was placed on studying the regulatory environment, the national and state day care system operation, as well as the need for a fee system to efficiently allocate resources. Rising costs and user demands for this service are presently increasing the need for additional funds to expand service availability throughout the State of Georgia and the U.S.

Two reports specifically address the problem of designing and evaluating sliding fee systems. Lee Cash completed a thesis at Georgia State University, in 1976, entitled "A Model for Sliding Fee Systems in Public Day Care". The objectives of the thesis were "to develop a predictive model that can describe the potential impacts of sliding fee schedules" and "to consider how a fee schedule should be selected and how the administrative machinery required to implement the proposed fee

schedule should be used to evaluate the impact and effectiveness of that schedule"[pg. 22-23]. The model developed has several limitations:

(1) The number of children in a family are not considered.

(2) The costs of day care are held constant.

(3) The model assumes that 11% of all 0-6 year olds at each income level will be served.

The output of Cash's model gives the number of children served, monthly fees, and monthly cost to the government over several income ranges for varying fee rates on a one time basis. This model provides one estimate at a single point in time. Cash concludes that there is a "lack of a clear methodology for developing sliding fee systems for public day care" and there is "a gap in the professional literature on the public administration of day care"[pg. 42].

Abt Associates, Inc., Cambridge, Massachusetts, published in 1972 a report "Sliding Fee Schedules-A Simulation Analysis of Child Care Service and Cost under Welfare Reform" for the Department of Health, Education and Welfare. The objective of Abt was "to identify a range of alternative fee schedules and to analyze the implications of each fee schedule in terms of the numbers and characteristics of families and

children likely to be served, the cost of child care services to the government, and the cost of participating families" [pg.I-1] . The report discusses in detail the payment mechanisms for subsidizing child care under H.R.1. Abt compares straight line and various types of curved fee schedules as they impact upward mobile families across the income breakeven point. They conclude that "the overall impact of curving fee schedules cannot be predicted exactly since static and dynamic effects may be very different" [pg.II-24] . The Abt model (called SIMFARE-C) does not consider income notch effects produced by loosing Medicaid, public housing, Aid to Families with Dependent Children, FICA, and federal and state income tax. The model does not take into account the IRS income tax deduction for child care. The model considers the manipulation of a single input data base at a one-time iteration, assuming everything remains constant throughout a one-year period. There is no consideration of the dynamic nature of the population environment over time. Abt assumed a perfectly elastic supply curve; in other words the supply is assumed to meet demand. Another significant drawback of the SIMFARE-C model is its great size and complexity. Abt states that the interpretation of "the results of any single run or pair of runs is very difficult without analysis of many other runs"

[pg. III-3].

The model produces a Summary Report with 19 key impact measures and 13 key program parameters. Detail Report #1 shows how the total number of families and children served are distributed among the 16 categories (by age and number of children). Detail Report #2 shows average government cost per child and per family. Detail Report #3 accounts for non-participating families in subsidized child care.

Presently, Abt Associates, Inc. is preparing a "National Day Care Center Supply Study" for the Department of Health, Education and Welfare. Their purpose is to update nationwide information on the characteristics of licensed day care centers. Abt proposes to provide a statistical profile of day care center industry by state. The profile will include:

(1) general characteristics of day care centers

(2) geographic distribution of centers

(3) distribution of staff characteristics

(4) financial characteristics

(5) characteristics of families served, including income, number of children enrolled per family and percentage of mothers working [pg. 3]

Abt proposes to design a model " to forecast the impact of alternative licensing regulations, alternative monitoring and funding strategies and variations in local conditions" on the following variables:[,pg. 4]

(1) licensed capacity

(2) enrollemnt

(3) number of paid staff and average wage

(4) average cost of child care per child-year

(5) average parent fee

(6) total private-paid enrollment

(7) numbers of profit and non-profit centers

The Abt supply study does not apply to the development of sliding fee systems. They plan to study available facilities and staff and not how a fee system would impact the day care system. In addition, Abt's scope is a national one, encompassing a large number of interrelated variables that might not be relevant in a state level child care system. Their supply data is to be completed in Winter 1978.

Summarizing, the literature research reveals very limited professional articles written on sliding fee systems for day care services. Only two reports specifically address sliding fee systems design. Lee Cash's thesis addresses the design of sliding fee systems. The model is limited by the assumption of a static environment and many factors are held constant. Abt Associates under the auspices of the Department of Health, Education and Welfare made a study on sliding fee system on a national scale. The large number of interrelated variables (about 400) limits the practicability and evaluation of the results. Abt and Cash do not consider a very important factor which is the dynamic nature of the day care environment over time.

The Abt Day Care Supply Study focuses on the national level. Its purpose is to update nationwide information on day care facilities, staff and programs. Hence, its study will gather information on a national basis and does not apply to the design and evaluation of sliding fee systems.

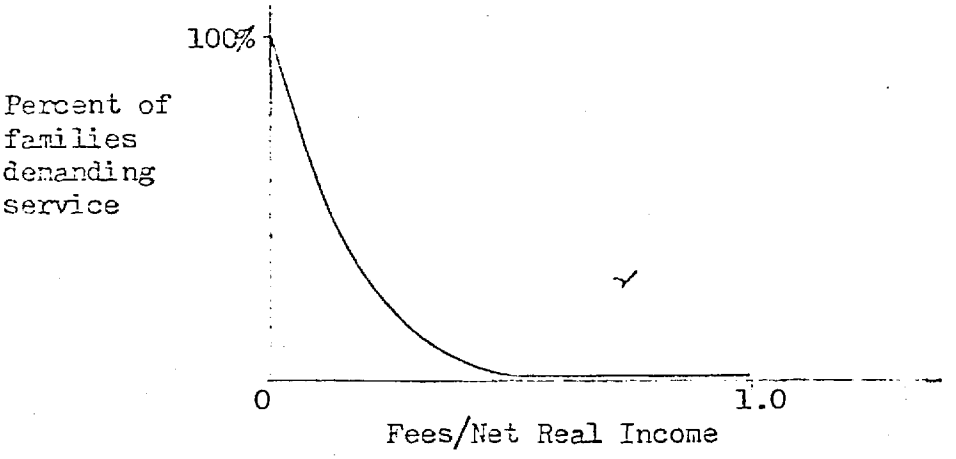
1.5 Methodology

The model will be written in FORTRAN, so that it can be widely interpreted and run on as many types of computers as possible. It will be a deterministic simulation and will employ fixed-interval (six months) clock updating. The maximum recommended run length will be five to ten years.

The major input variables will include:

1. An initial matrix of the numbers of families cross-classified by income ranges and family size (one matrix for each of the ten geographic DHR Districts).
2. Factors by which families are expected to move to adjacent cells in the family income/size matrices during each six-month update.
3. The distribution of children by age (i.e., infant, pre-school, elementary school) for each DHR District.
4. An assumed demand curve for child care services expressed in terms of the percentage of families demanding the service as a decreasing function of fees as a percentage of net real income (adjusted for loss of other benefits and increasing taxes). A typical illustration of a demand curve is shown

below.



5. The inital number of child care slots available by DHR District, and assumptions as to how that supply will increase over time. Four possible assumptions for the increase in supply are:
   a. The supply will increase linearly at some specified slope.
   b. The supply will increase exponentially by some specified percentage in each six-month period.
   c. The supply will increase in response to demand but with a specified time lag.
   d. The supply will increase as a function of income received from fees.
6. Cost per child to the state of providing day care services as a function of child age and DHR District. Also, the expected rate at which these costs will inflate over time.
7. Typical family budgets (i.e., income allocations) for urban and rural families.
8. Schedules of benefits lost (e.g., food stamps, public housing, Aid to Families with Dependent Children, etc.) and additional taxes incurred as family income rises.

9. The sliding fee scale itself, expressed in terms of fees charged as a function of family income adjusted for family size.

During each six-month update of a simulation run, the following sequence of computational operations will be performed:

1. Revise the family income/size matrices for each of the DHR Districts.
2. Determine the demand for day care services by family income range and child age in each district.
3. Update the current supply of child care slots in each district.
4. Determine the total cost of providing the service, the fees paid by participating families, and the cost to the government.

The results of the above computations will be printed at the end of each update. In addition, appropriate cumulative results will be printed at the end of each run.

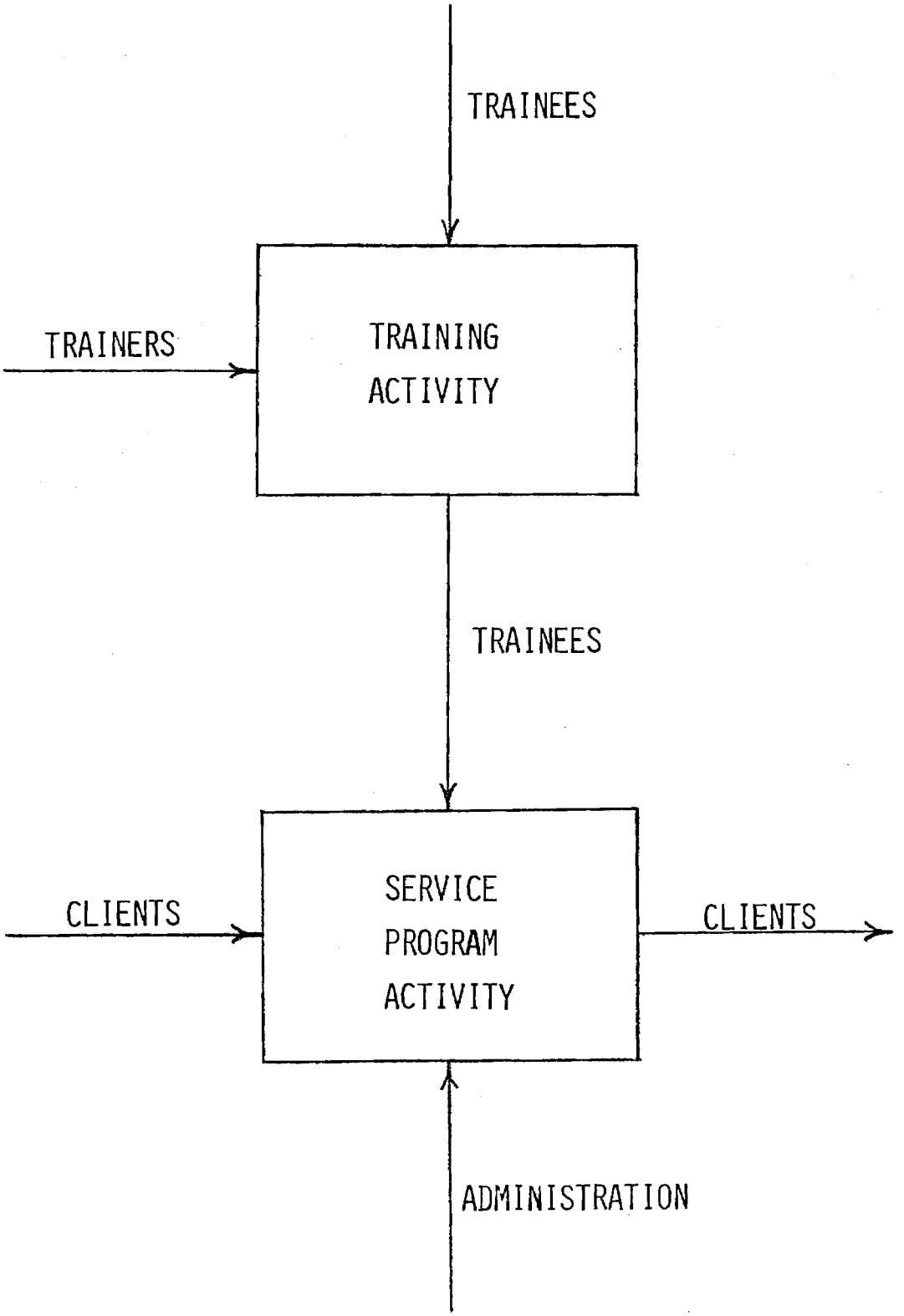
The primary method for validating the model will be to run it under various extreme input conditions for which the appropriate results can be logically predicted. As with any model, validity will be a function of both the structure of the model, and the parameters and input data on which it operates. It is re-emphasized that much of the required data is presently non-existent, and the responsibility for developing such data in the future can rest only with the user on a continuing basis. Therefore, the primary focus of the model validation

process will be on the model's structure and computational procedures..

In addition to validation runs, up to five test runs will be made using the 1970 Census Bureau data.

APPENDIX E

Transparencies
used in the
Title XX Training Contractors' Meeting
March 17, 1977


TRAINEES
TRAINERS
TRAINING ACTIVITY
TRAINEES
CLIENTS
SERVICE PROGRAM ACTIVITY
CLIENTS
ADMINISTRATION

TYPES OF EVALUATIONS

1. PROCESS/RESOURCE EVALUATIONS

   - LOOK AT WHAT WE ARE DOING.

   - LOOK AT THE RESOURCES WE ARE USING.

   - LOOK AT HOW MUCH THE TRAINEES ENJOY AND "FEEL GOOD" ABOUT THE TRAINING EXPERIENCE.

2. RESULTS, IMPACT, OUTCOME EVALUATIONS

   - LOOK AT THE IMPROVEMENT IN THE KNOWLEDGE, SKILLS, ATTITUDES, ETC. OF THE TRAINEES.

   - LOOK AT THE IMPROVEMENT IN THE TRAINEES' PERFORMANCE, BEHAVIOR, ETC. ON THE JOB.

   - LOOK AT THE IMPROVEMENT IN THE PERFORMANCE OF PROGRAMS IN WHICH THE TRAINEES WORK.

REQUIREMENTS OF IMPACT EVALUATIONS

1. SHOW THAT A CHANGE OCCURRED.

2. SHOW THAT THE CHANGE WAS CAUSED PRIMARILY BY THE TRAINING.

THE ABOVE REQUIREMENTS IMPLY THE NEED FOR:

- EXPERIMENTAL DESIGN
- MEASUREMENT
- STATISTICAL ANALYSIS

WHY EVALUATE TRAINING PROGRAMS

1. We are required to do so.

2. Evaluation provides input to decision making.

   - Within the program

   - Above the program

3. Evaluation builds knowledge.

THE BURDEN OF PROOF

- ALTERNATIVE PRESUMPTIONS -

- ASSUME SUCCESS UNLESS FAILURE IS PROVEN.

- ASSUME FAILURE UNLESS SUCCESS IS PROVEN.

A MIDDLE GROUND -

ASSUME IGNORANCE UNTIL INFORMATION IS AVAILABLE.

BARRIERS TO EVALUATION

- Conceptual Problems
- Technical/Methodological Problems
- Organizational/Political/Behavioral Problems

EXCUSES FOR FAILING TO PERFORM IMPACT EVALUATIONS

1. OUR PROGRAM IS UNIQUE, AND THE RESULTS CANNOT BE MEASURED.
2. WE SHOULD DEVOTE OUR TIME AND EFFORT TO TRAINING.
3. EVALUATION MEASUREMENTS MIGHT CREATE STRESS FOR THE TRAINEES.
4. THINGS ARE CLEARLY GOING WELL, SO THAT A RIGOROUS EVALUATION IS NOT NECESSARY.

CONTINUOUS VS. AFTER-THE-FACT EVALUATIONS

ADVANTAGES OF CONTINUOUS EVALUATIONS:

1. ENCOURAGES CAREFUL ANALYSIS OF PROGRAM OBJECTIVES BEFORE BECOMING CAUGHT UP IN ACTIVITY. MEASUREMENT OF STARTING POINT DISCOURAGES OVER-PROMISING.

2. CAN SHOW THAT CHANGES OCCUR.

3. CAN USE EVALUATION INFORMATION IN MANAGING THE PROGRAM.

4. PROVIDES A SENSE OF THE DESIRED RATE OF PROGRESS. SHIFTS PRESSURE TO THE FRONT END OF THE PROGRAM.

5. MEASURABLE PROGRESS YIELDS SATISFACTION. EVALUATION CAN BE MOTIVATING.

6. CAN IMPROVE EVALUATION PROCEDURES BASED ON EXPERIENCE.

EXPERIMENTAL DESIGNS FOR TRAINING PROGRAM EVALUATION

1. PROVING THAT A CHANGE OCCURRED.

MEASURE	TRAIN	MEASURE

2. CONTROLLING FOR "HISTORY" AND "SUBJECT MATURATION."

MEASURE	TRAIN	MEASURE
MEASURE	-	MEASURE

3. CONTROLLING FOR EFFECTS OF PRETESTING AND PRETEST/TRAINING INTERACTION.

MEASURE	TRAIN	MEASURE
MEASURE	-	MEASURE
-	TRAIN	MEASURE
-	-	MEASURE

4. Controlling for ego involvement and "Hawthorne Effect." Testing for cost effectiveness.

Measure	Train	Measure
Measure	Placebo	Measure
Measure	-	Measure

5. Controlling for delay or decay in effects.

Measure	Train	M M M M

6. A powerful hybrid design.

Measure	Train	M M M M
-	Train	M M M M
Measure	Placebo	M M M M

MEASUREMENT SCALES

1. NOMINAL

   - SERVES ONLY TO DIFFERENTIATE OR CLASSIFY.

2. ORDINAL

   - INDICATES ORDER OF RELATIVE MAGNITUDE OR RANK.

   - INTERVALS BETWEEN ADJACENT UNITS NOT EQUAL IN GENERAL.

3. INTERVAL

   - EQUAL INTERVALS BETWEEN ADJACENT UNITS.

   - ARBITRARY ZERO POINT.

4. RATIO

   - ABSOLUTE ZERO POINT.

QUALITIES OF MEASUREMENTS

1. Accuracy, Validity

   - Is the measurement correct (or correct on the average)?

   - Affected by systematic errors or biases.

   - Errors corrected by:

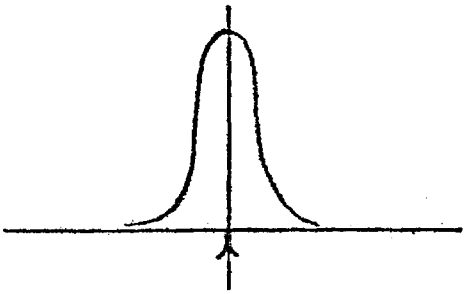
      A. Improving measurement procedures to eliminate bias.

      B. Adjusting measurement results.

2. Precision, Reliability, Consistency

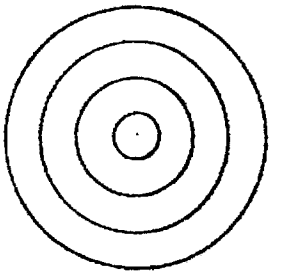
   - Do repeated measurements give the same results?

   - Affected by random errors.

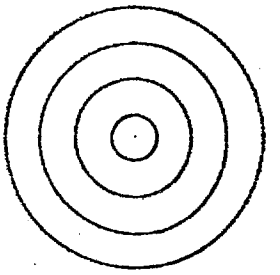


CORRECT VALUE

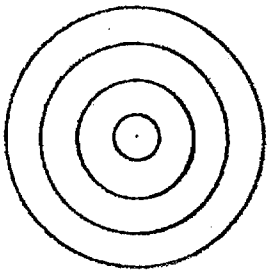
ACCURATE & PRECISE

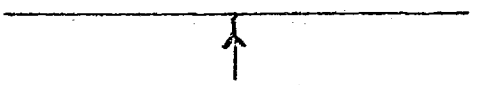


ACCURATE BUT NOT PRECISE

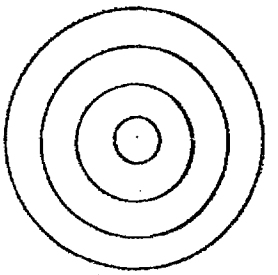


PRECISE BUT NOT ACCURATE





NEITHER ACCURATE NOR PRECISE



SOURCES OF MEASUREMENT ERRORS

1. CONDITIONS OF MEASUREMENT

2. MEASUREMENT INSTRUMENT

3. INDIVIDUAL PERFORMING THE MEASUREMENT

4. SUBJECT OF MEASUREMENT

5. SAMPLING PROBLEMS

   - SAMPLE SIZE

   - SAMPLE BIAS

APPENDIX F

Post-test

COMPREHENSIVE EXAM
PROGRAM EVALUATION METHODOLOGIES

NAME: _____

NOTE: This is a closed-book, closed-notes exam. Please complete all studying that you wish to do before you begin the exam. Please do not talk with others about the exam after you have begun working on it. Your completed exam should be mailed no later than 21 January to:

Dr. Thomas B. Clark
School of Industrial & Systems Engineering
Georgia Tech
Atlanta, GA 30332

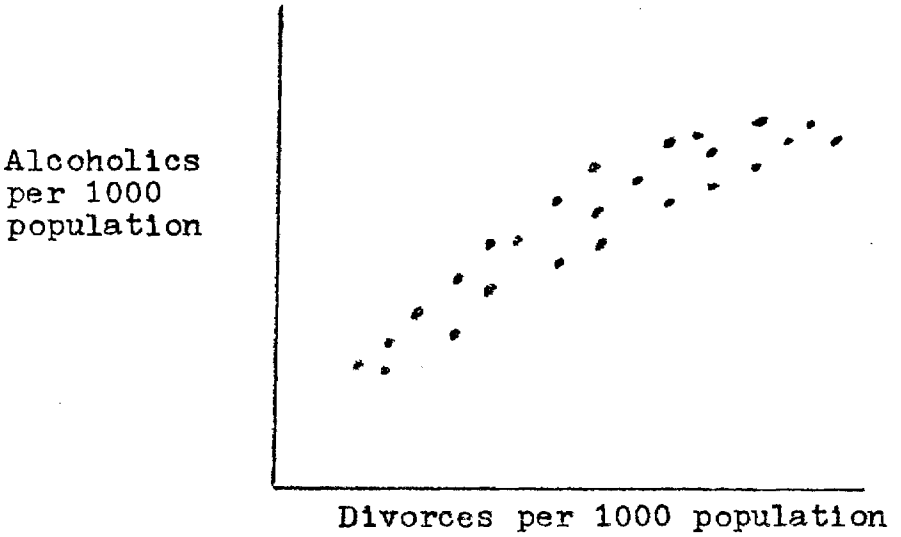
(2 pts.) 1. Briefly explain the difference between "process evaluation" and "impact (or results, outcome) evaluation."

(2 pts.) 2. Briefly distinguish between "efficiency" and "effectiveness" as criteria for program evaluation.

(2 pts.) 3. Other than the fact that program evaluations are now required for Federally funded programs, for what major reasons should social service programs be continuously evaluated?

(4 pts.) 4. Briefly explain the concepts of "internal validity" and "external validity" as they relate to evaluation research.

(5 pts.) 5. The graph shown below is the result of a research study seeking to determine the causes of alcoholism in the U.S. The data cover the last 25 years. Each point on the graph shows the number of known alcoholics per 1000 population for a given year versus the number of divorces obtained per 1000 population in that same year. Give <u>three</u> possible explanations for the relationship shown between the two variables.



(4 pts.) 6. Explain the "Hawthorne Effect" and describe a research design that allows the effect to be measured.

(3 pts.) 7. Explain the "regression effect" as it could relate to evaluation research.

(4 pts.) 8. Explain the procedure and purpose of the "Solomon 4-group" research design.

(5 pts.) 9. For each of the following measurements, indicate whether an interval, ordinal, ratio, or nominal scale would be involved by placing an I, O, R, or N in the space provided:

____Items on an instructor evaluation form on which you rate various aspects of the instructor's performance by checking "outstanding," "good," "fair," or "poor."

____The numbers 1, 2, 3, 4, and 5 which are assigned to different child care centers within a regional program to provide an easy means of identification.

____The years 1972, 1970, 1975, 1975, and 1973 in which each of the above child care centers were opened.

____The number of years that each of the above child care centers have been in operation (i.e., 4, 6, 1, 1, 3).

____A ranking of all the individuals participating in a given program in order of their annual income, where the #1 ranked individual has the lowest income.

(3 pts.) 10. Indicate whether each of the following statements is true or false by circling T or F:

T F It is appropriate to compute averages of measurements on an <u>interval</u> scale.

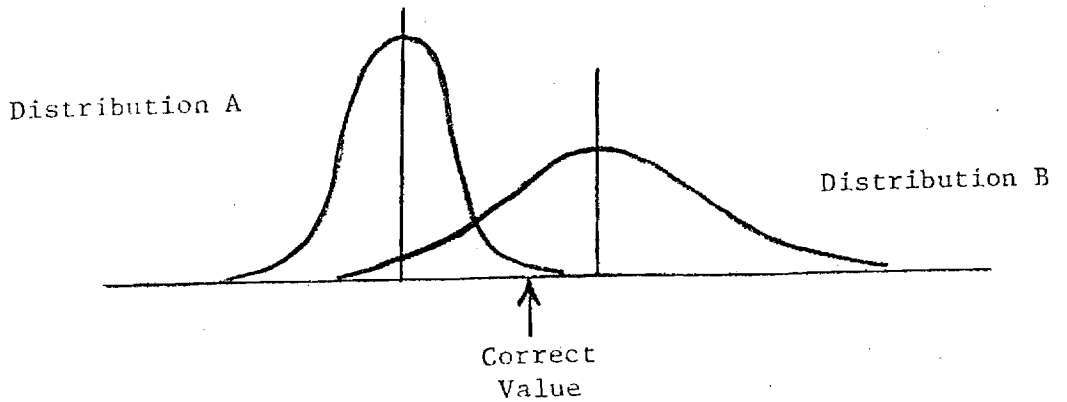
T F It is appropriate to add and subtract measurements on an <u>ordinal</u> scale.

T F It is appropriate to form ratios of measurements on an <u>interval</u> scale.

(4 pts.) 11. Two distributions of measurements of the same item are shown below along with the correct value for the measurement.

a. Which distribution is more precise? _____

b. Which distribution is more accurate? _____



(3 pts.) 12. We have identified several possible sources of measurement errors, one of which is the person who takes or records the measurement. Identify <u>three</u> other possible sources.

(3 pts.) 13. Contrast the nature of measurement errors that cause inaccuracy with the nature of errors that cause imprecision.

(4 pts.) 14. Explain "reliability" and "validity" as characteristics of survey measurements.

(3 pts.) 15. a. <u>Identify</u> <u>three</u> well-known attitude measurement scales.

(4 pts.) b. <u>Describe</u> <u>any</u> <u>two</u> of these attitude measurement scales. (Continue your answer on the back of this page if necessary.)

16. Briefly describe each of the following sampling procedures:

(2 pts.) a. Stratefied Sampling

(2 pts.) b. Systematic Sampling

(2 pts.) c. Multistage Cluster Sampling (unstratefied)

(4 pts.) 17. Explain the problem of "nonrespondent bias."

(4 pts.) 18. a. You want to know what proportion of currently licensed automobile drivers in Georgia had a complete physical examination during the period 1/1/74 to 12/31/76. You plan to conduct a telephone survey of a randomly selected sample of drivers. In an initial survey of 60 drivers, 36 said that they had had a physical exam during the specified period. How large a sample would you need to take in order to be 95% certain that the proportion you obtained would be within $\pm 2\frac{1}{2}\%$ of the proportion you would obtain if you asked all licensed drivers in Georgia? Set up the calculation only; do not carry out the arithmetic.

$$N = \frac{Z^2 (p)(1-p)}{E^2}$$

Z	% of Values in Normal Distributions Contained Within $\mu \pm Z \cdot \sigma$
0.67	50.0%
1.00	68.3
1.65	90.0
1.96	95.0
2.00	95.5
2.58	99.0
3.00	99.7

(2 pts.) b. What can you say concerning the accuracy of the proportion that you would obtain from the above sample?

(4 pts.) 19. Explain the nature of "Type I (or alpha) errors" and "Type II (or beta) errors" in statistical hypothesis tests.

(4 pts.) 20. Contrast the circumstances in which "analysis of variance" would be an appropriate statistical procedure versus circumstances in which "correlation analysis" would be more appropriate.

(8 pts.) 21. Indicate whether each of the following statements concerning nonparametric statistical procedures is true or false by circling T or F:

T F They are generally more powerful than parametric procedures.

T F They can be used in situations where both the independent variables and the dependent variables are measured on nominal or ordinal scales.

T F They are generally less difficult to understand and use than parametric procedures.

T F They generally involve less restrictive technical assumptions than parametric procedures.

(2 pts.) 22. What is SPSS?

(3 pts.) 23. What major steps are required in the development of a new computer program?

(6 pts.) 24. a. Construct a CPM network of the project described below. Perform all forward, reverse, and total slack computations, recording the results clearly on the network.

Activity	Preceding Activities	Duration
A	-	5 days
B	A	4 "
C	A	3 "
D	A	7 "
E	B,D	8 "
F	C	10 "
G	C,D	8 "
H	E,F,G	3 "

(2 pts.) b. What sequence(s) of activities comprise the critical path(s) in the above project?

* * * END * * *